PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/10 | A3

(11) International Publication Number: **WO 99/65927**

(43) International Publication Date: 23 December 1999 (23.12.99)

(21) International Application Number: PCT/US99/13479

(22) International Filing Date: 15 June 1999 (15.06.99)

(30) Priority Data:
60/089,649 17 June 1998 (17.06.98) US

(71) Applicant: MAXYGEN, INC. [US/US]; 515 Galveston Drive, Redwood City, CA 94063 (US).

(72) Inventor: KREBBER, Claus, M.; 1935 Rock Street #1, Mountain View, CA 94043 (US).

(74) Agents: APPLE, Randolph, T. et al.; Townsend and Townsend and Crew LLP, 8th floor, Two Embarcadero Center, San Francisco, CA 94111 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report:
10 February 2000 (10.02.00)

(54) Title: METHOD FOR PRODUCING POLYNUCLEOTIDES WITH DESIRED PROPERTIES

(57) Abstract

The invention provides methods for the production of polynucleotides with a desired property (e.g., conferring a desired phenotype and/or encoding polypeptide with a desired property) which is selectable or can be screened for. The method includes making insertions and/or deletions at random sites in DNA segments in a population. In some embodiments the random insertions and deletions are made recursively.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/10 | A3

(11) International Publication Number: **WO 99/65927**

(43) International Publication Date: 23 December 1999 (23.12.99)

(21) International Application Number: PCT/US99/13479

(22) International Filing Date: 15 June 1999 (15.06.99)

(30) Priority Data:
60/089,649 17 June 1998 (17.06.98) US

(71) Applicant: MAXYGEN, INC. [US/US]; 515 Galveston Drive, Redwood City, CA 94063 (US).

(72) Inventor: KREBBER, Claus, M.; 1935 Rock Street #1, Mountain View, CA 94043 (US).

(74) Agents: APPLE, Randolph, T. et al.; Townsend and Townsend and Crew LLP, 8th floor, Two Embarcadero Center, San Francisco, CA 94111 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report:
10 February 2000 (10.02.00)

(54) Title: METHOD FOR PRODUCING POLYNUCLEOTIDES WITH DESIRED PROPERTIES

(57) Abstract

The invention provides methods for the production of polynucleotides with a desired property (e.g., conferring a desired phenotype and/or encoding polypeptide with a desired property) which is selectable or can be screened for. The method includes making insertions and/or deletions at random sites in DNA segments in a population. In some embodiments the random insertions and deletions are made recursively.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

INTERNATIONAL SEARCH REPORT

Int ional Application No
PCT/US 99/13479

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/10

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 285 123 A (SOUMEN SOKERY OY) 5 October 1988 (1988-10-05) * claims; pages 1-8 * ---	1-25
X	WO 92 18645 A (DIAGEN) 29 October 1992 (1992-10-29) * whole disclosure * ---	1-25
X	WO 97 35966 A (MAXYGEN) 2 October 1997 (1997-10-02) * claims; pages 1,3,7,8 * ---	1-25
X	HANES, J. & PLÜCKTHUN, A.: "In vitro selection and evolution ..." PROC. NATL. ACAD. SCI. USA, vol. 94, May 1997 (1997-05), pages 4937-4942, XP002077149 * whole disclosure * ---	1-25
	-/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
14 December 1999	21/12/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Hermann, R

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Int ional Application No
PCT/US 99/13479

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	STEMMER, W.P.C.: "Rapid evolution of a protein in vitro ..." NATURE, vol. 370, August 1994 (1994-08), pages 389-391, XP002912216 * whole disclosure * -----	1-25

1

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inte ional Application No
PCT/US 99/13479

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 285123 A	05-10-1988	FI 881530 A	04-10-1988
		JP 1020089 A	24-01-1989
WO 9218645 A	29-10-1992	DE 4112440 C	22-10-1992
		AT 171479 T	15-10-1998
		DE 59209506 D	29-10-1998
		EP 0583265 A	23-02-1994
		JP 6506355 T	21-07-1994
		US 5795747 A	18-08-1998
WO 9735966 A	02-10-1997	US 5837458 A	17-11-1998
		AU 2337797 A	17-10-1997
		AU 2542697 A	17-10-1997
		CA 2247930 A	02-10-1997
		EP 0932670 A	04-08-1999
		EP 0906418 A	07-04-1999
		WO 9735957 A	02-10-1998
		AU 1087397 A	19-06-1997
		CA 2239099 A	05-06-1997
		EP 0876509 A	11-11-1998
		EP 0911396 A	28-04-1999
		WO 9720078 A	05-06-1997

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07H 21/00

A2

(11) International Publication Number: **WO 99/65927**

(43) International Publication Date: 23 December 1999 (23.12.99)

(21) International Application Number: PCT/US99/13479

(22) International Filing Date: 15 June 1999 (15.06.99)

(30) Priority Data:
60/089,649 17 June 1998 (17.06.98) US

(71) Applicant: MAXYGEN, INC. [US/US]; 515 Galveston Drive, Redwood City, CA 94063 (US).

(72) Inventor: KREBBER, Claus, M.; 1935 Rock Street #1, Mountain View, CA 94043 (US).

(74) Agents: APPLE, Randolph, T. et al.; Townsend and Townsend and Crew LLP, 8th floor, Two Embarcadero Center, San Francisco, CA 94111 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: METHOD FOR PRODUCING POLYNUCLEOTIDES WITH DESIRED PROPERTIES

(57) Abstract

The invention provides methods for the production of polynucleotides with a desired property (e.g., conferring a desired phenotype and/or encoding polypeptide with a desired property) which is selectable or can be screened for. The method includes making insertions and/or deletions at random sites in DNA segments in a population. In some embodiments the random insertions and deletions are made recursively.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

METHOD FOR PRODUCING POLYNUCLEOTIDES WITH DESIRED PROPERTIES

Field of the Invention

The present invention relates to methods for the production of polynucleotides conferring a desired phenotype and/or encoding a polypeptide having an advantageous predetermined property which is selectable or can be screened for.

Background of the Invention

Traditional molecular biological methods for generating novel genes and proteins generally involved rational or directed mutation. An example is the generation of a polynucleotide encoding a fusion or chimeric protein by using known restriction sites to combine functional domains from two characterized proteins. Another example is the introduction of a point mutation at a specific site in a polypeptide. Although useful, the power of these and similar methods is limited by the requirement for sequence or restriction map information to facilitate the mutagenesis, and by the limited number of variants that can be efficiently generated.

An alternative approach to the generation of variants uses random recombination techniques such as "DNA shuffling" (Patten et al., 1997, *Curr. Opin. Biotech.* 18:724-733). DNA shuffling entails performing iterative cycles of recombination and screening or selection to "evolve" individual genes, whole plasmids or viruses, multigene clusters, or whole genomes. Such techniques do not require the extensive analysis and computation required by conventional methods for engineering of polynucleotides and polypeptides. Moreover, DNA shuffling allows the recombination of large numbers of mutations in a minimum number of selection cycles, in contrast to traditional, pairwise recombination events. Thus, DNA shuffling techniques provide advantages in that they provide recombination between mutations in any or all of these, thereby providing a very fast way of exploring the manner in which different combinations of mutations can affect a desired result.

The present invention provides methods that may be used alone or in combination with random recombination techniques such as DNA shuffling to generate novel polynucleotides having, or encoding a polypeptide having, a desired property or combination of properties.

Summary of the Invention

In one aspect, the invention provides a method of producing a DNA segment having a desired property or combination of properties by mutating a substrate population. The method involves:

a) mutating a substrate population that includes a plurality of DNA segments by:

i) making insertions at random sites in the segments (random insertion),

ii) making deletions at random sites in the segments (random deletion), or both, to produce a mutated population including mutated DNA segments,

b) screening the mutated population to obtain a first selected population that includes at least one DNA segment with a first desired property,

c) mutating the first selected population by making random insertions, random deletions, or both, to produce a recursively mutated population, and,

d) screening the recursively mutated population to obtain a recursively selected population that includes at least one DNA segment with a second desired property.

In some embodiments the method further includes at least one additional cycle of mutation and screening (e.g., mutating the recursively selected population and screening the resulting recursively mutated population to obtain new recursively selected population with a desired property) after step (d). In some embodiments, shuffling of one or a combination of polynucleotides in a recursively selected population is carried out.

In various embodiments, the second desired property may be the same or different from the first desired property, and may be a combination of properties. In some embodiments, the polynucleotides in the recursively selected population have a property that is enhanced when compared to the polynucleotides in the first selected population. In some embodiments the substrate population includes DNA segments encoding a polypeptide, a catalytic RNA, a promoter sequence or a vector. In some embodiments the substrate population is homogeneous. In some embodiments a polynucleotide that encodes a polypeptide is screened for an activity such as an enzymatic activity, a substrate specificity, or a binding activity of a polypeptide.

In another aspect, the invention provides a method of producing a DNA segment having a desired property by:

a) mutating a first substrate population that includes a plurality of DNA segments by:

i) making insertions at random sites in the segments (random insertion),

ii) making deletions at random sites in the segments (random deletion), or both, to produce a first mutated population of mutated DNA segments;

b) mutating a second substrate population that includes a plurality of DNA segments by:

i) making insertions at random sites in the segments,

ii) making deletions at random sites in the segments, or both

to produce a second mutated population of mutated DNA segments;

c) recombining the first substrate population and the second substrate population to produce a recombined population; and,

d) screening the recombined population to identify at least one DNA segment with the desired property.

In one embodiment, the first and second mutated populations are screened to produce a first and second selected population, each having a desired property, and the selected populations are recombined.

In various embodiments, the recombination may be achieved by shuffling or directed recombination. In some embodiments the first desired property and the second desired property are the same. In some embodiments the substrate population includes DNA segments encoding a polypeptide, a catalytic RNA, a promoter sequence or a vector. In some embodiments the substrate population is homogeneous. In some embodiments a polynucleotide that encodes a polypeptide is screened for an activity such as an enzymatic activity, a substrate specificity, or a binding activity of a polypeptide.

In another aspect, the invention provides a method of producing a DNA segment having a desired property by:

a) mutating a substrate population that includes a plurality of DNA segments by:

i) making insertions at random sites in the segments,

ii) making deletions at random sites in the segments;

or both, to produce a mutated population of mutated DNA segments;

b) screening the mutated population to obtain a selected population that includes at least one DNA segment with the desired property;

c) shuffling at least one DNA segment for the selected population to produce a recombined population;

d) screening the recombined population for a desired property.

In one embodiment, the shuffling involves conducting a polynucleotide amplification process on overlapping segments of at least one polynucleotide from the selected population under conditions under which one segment serves as a template for extension of another segment, to generate a population of recombinant polynucleotides.

In some embodiments the substrate population includes DNA segments encoding a polypeptide, a catalytic RNA, a promoter sequence or a vectors. In some embodiments the substrate population is homogeneous. In some embodiments a polynucleotide that encodes a polypeptide is screened for an activity such as an enzymatic activity, a substrate specificity, or a binding activity of a polypeptide.

Brief Description of the Figures

Figure 1 provides a flow-diagram of an embodiment of the invention in which recursive steps of random insertion or deletion and screening are employed to produce a DNA segment with a desired property.

Figure 2 provides a flow-diagram of an embodiment of the invention in which random insertion or deletion is carried out on two different substrate populations, which are then recombined.

Figure 3 provides a flow-diagram of an embodiment of the invention in which random insertion or deletion, screening, and random recombination steps are employed to produce a DNA segment with a desired property.

Detailed Description

I. Definitions

The following terms are defined to provide additional guidance to one of skill in the practice of the invention:

The term "shuffling," as used herein, refers to techniques for random recombination between substantially homologous but non-identical polynucleotides. Various shuffling methods are described in Patten et al., 1997, *Curr. Opin. Biotech.* 8:724-733; Stemmer, 1994, *Nature* 370:389-391; Stemmer et al., 1994, *Proc. Natl. Acad. Sci. USA* 91:10747-10751; Zhao et al., 1997, *Nucleic Acids Res.* 25:1307-1308; Crameri et al.,1998 , *Nature* 391: 288-291; Crameri et al., 1997, *Nat. Biotech.* 15:436-438; Arnold et al., 1997, *Adv. Biochem. Eng. Biotechnol.* 58:2-14; Zhang et al., 1997, *Proc. Natl. Acad. Sci. USA* 94:4504-4509; Crameri et

al., 1996, *Nat. Biotechnol.* 14:315-319; Crameri et al., 1996, *Nat. Med.* 2:100-102; PCT publications WO95/22625; WO97/20078; WO97/35957; WO97/35966; WO98/13487; WO98/13485; PCT 98/00852; PCT 97/24239, and references therein. Shuffling techniques are also described in the following U.S. patents and patent applications: U.S. Patent No. 5,605,793; U.S. Patent Applications Serial Nos: 08/537,874; 08/621,859; 08/792,409; 08/769,062; 08/822,589; 09/021,769; 60/074,294; 08/722,660; 08/938,690. Each of the aforementioned patents, applications, and publications is incorporated herein by reference in its entirety and for all purposes. One method of shuffling comprises conducting a polynucleotide amplification process on overlapping segments of a population of variants of a polynucleotide under conditions whereby one segment serves as a template for extension of another segment, to generate a population of recombinant polynucleotides, and screening or selecting a recombinant polynucleotide or an expression product thereof for a desired property. Some methods of shuffling use random point mutations (typically introduced in a PCR amplification step) as a source of diversity.

The term "oligonucleotide, " as used herein, generally refers to polynucleotides shorter than about 50 bases (e.g., about 6, 9, 12, 15, 18, 21, 25, 35, or 50 bases in length). The term "polynucleotide," as used herein, refers to both oligonucleotides and longer molecules (e.g., at least about 60, 100, 200, 300, 500, 1000, 5000, 10,000 bases or base pairs in length, or even longer. The oligo and polynucleotides used in the present invention are usually DNA molecules, and typically are double stranded.

The term "property," as used herein, refers to any characteristic or attribute of a polynucleotide (or, e.g., an encoded polypeptide or RNA) that can be selected for or detected in a screening system, including, for example, enzymatic or binding activity of a polynucleotide or an encoded polypeptide (e.g., a new activity or enhanced or diminished level of a preexisting activity), fluorescence, properties conferred on a cell comprising a particular polynucleotide, a binding activity (e.g., the property of binding, or being bound by, a specific target molecule, such as receptor, ligand, antibody or antibody fragment, antigen, epitope, or other biological macromolecule). The property may be an attribute of a sequence controlling transcription (e.g., promoter strength, regulation), a sequence affecting RNA processing (e.g., RNA stability or splicing), a sequence affecting translation (e.g., level, regulation, post-transcriptional modification), or a sequence affecting other expression property of a gene or transgene; a replicative element, a protein-binding element; a vector; an encoded protein (e.g., enzymatic

activity and specificity, binding activity and specificity, pI, stability to denaturation), an encoded RNA (e.g., mRNA or catalytic RNA), and the like. Additional examples are described herein or in the references incorporated herein, or will be apparent to one of skill upon reading this disclosure.

The term "evolve," as used herein, refers to the process of introducing variation into a population of macromolecules and selecting or screening for acquisition of a desired property or the partial acquisition of a desired property, resulting in the generation of one or more molecules different from the molecules of the starting population.

II. Overview

The present invention provides novel methods for the generation of polynucleotides having a desired property (e.g., an advantageous predetermined property which is selectable or can be screened for). In one aspect, the invention provides methods for generating diversity in a population of polynucleotides by random insertion or deletion of sequences and identification of variants with new or enhanced properties. In some embodiments, multiple cycles of insertion/deletion and screening are carried out. In some embodiments, the properties of the variants are evolved by one or more of a variety of methods.

Typically the mutated polynucleotides are double stranded DNA segments. Examples of suitable DNA segments include DNAs comprising genes, gene fragments, groups of genes, vectors, polypeptide-coding sequences, expression regulatory sequences (e.g., promoters, enhancers), and the like.

In one embodiment of the invention, a population of polynucleotides (i.e., a substrate population) is mutated by random insertion or deletion, and the resulting mutated population is screened to identify a subpopulation of species with a desired property (i.e., a selected population). The selected population is then itself mutated by random insertion or deletion, and the resulting twice mutated population is again subjected to screening to produce a new selected population. The second round of screening can be for the same or a similar property as screened for in the earlier round, or for an entirely different property. For example, when a substrate population of vectors is mutated, the first screen could be for species that have acquired a sequence conferring chloramphenicol resistance not found in the substrate population and the second screen could be for increased chloramphenicol resistance (the same or similar property), or, alternatively, in subsequent rounds of mutation and screening for the acquisition

of a sequence conferring tetracycline resistance (a different property). The process of mutation and selection can be carried out for multiple cycles, if desired, to generate one or more novel DNA segments that have a specific desired property or combination of properties. For example, in some embodiments at least 2, 5 or 10 cycles of random insertion/deletion and screening will be carried out. Following two or more cycles of mutation and selection, at least one polynucleotide species having the desired property or properties (e.g., an activity not found in the starting population of polynucleotides) is isolated from the subpopulation. This process is outlined generally in Fig. 1; however, the figure is presented solely to assist the reader and is not intended to limit the invention in any way.

In another embodiment, two or more different substrate populations are mutated by random insertion or deletion, producing corresponding mutated populations. In many embodiments, the two-or-more mutated populations are screened for particular desired properties (e.g., each mutated population is screened for a different property). Following production of the two or more mutated populations (or following screening if it takes place), polynucleotide segments from each of the mutated populations are recombined to produce a single recombined population. The recombination may be carried out by DNA shuffling, or, alternatively, using "classical" molecular cloning techniques in which a selected region in one population of polynucleotides is cloned into a specific site (e.g., a restriction site) in a second population of polynucleotides. "Classical" techniques include (i) restriction of two populations of DNA molecules and ligation of fragments from one of the populations into a restriction site in the DNA of the second population, (ii) amplification of a region of one polynucleotide population (e.g., by PCR or inverse PCR) and ligation into the polynucleotides of the second population, (iii) and other methods known in the art. The recombined population is then screened for the desired property(s). In some embodiments, subsequent cycles of random insertion/deletion or recombination and screening are carried out. This process is outlined in Fig. 2; like Figure 1, this figure is not intended to limit the invention.

In a third embodiment, a substrate population of polynucleotides is mutated by random insertion or deletion, the resulting mutated population is screened to identify species with a desired property (e.g., a "selected population"). The selected population (or a specie or species isolated from it) is then evolved by random recombination (including random recombination combined with point mutation), which may be recursive or single cycle random recombination. This process is outlined in Fig. 3; this figure also is not intended to limit the invention.

The invention will now be described in greater detail.

III. Mutating the Substrate Population

a) Generally

An initial step in the method of the invention is the introduction of insertions or deletions at random sites in a population of polynucleotides. Mutations and deletions are sometimes collectively referred to herein as "mutations." For convenience, a population of polynucleotides into which mutations are to be introduced may be referred to as the "substrate population."

Although the method can be carried out on any polynucleotides that can be mutated in a random fashion by insertion or deletion, as noted *supra* the polynucleotides will most often be DNA molecules (including cDNA), usually double-stranded DNA molecules. The DNA molecules making up the substrate population may be of any of several types, including DNA molecules comprising polypeptide coding sequences (e.g., encoding a protein, multiple proteins, or portions of a protein), regulatory DNAs (e.g., promoters, enhancers), vectors (e.g., an expression vector), and viruses (e.g., to produce attenuated virions). These DNA molecules are sometimes also referred to as "DNA segments."

The substrate population will comprise a plurality of DNA segments, typically at least 10^2, more often at least 10^4, or at least 10^6 DNA segments. In many embodiments, the DNA segments in any particular substrate population are identical to each other, being derived from a single parental DNA (e.g., plasmid DNAs prepared from the same bacterial culture). Such a population is a "homogeneous" substrate population. In some embodiments, however, the substrate population includes DNA segments that are not identical such as the following: DNA segments that differ from each other by point mutations (e.g., molecules that have been generated from a template using error-prone PCR) or other mutations (e.g., insertions or deletions); DNA segments that are related as homologs from different organisms; and DNA segments that are related to each other because they are products of DNA shuffling reactions (see, e.g., Patten et al., 1997, *Curr. Opin. Biotech.* 8:724). In a related embodiment, the substrate population will comprise DNA segments having unrelated sequences (for example, a substrate population comprising several different plasmid vectors), usually with a plurality (e.g., at least 10^2 or 10^6) of each species present.

Mutations (insertions or deletions or both) are introduced into the DNA segments

in the substrate population. For convenience, the population of polynucleotides that has been mutated may be referred to as the "mutated population." An important aspect of the present invention is that the mutations are introduced at random sites in the DNA segments. "Random," in this context, has its usual meaning and refers to insertions and deletions that (i) are not made at predetermined sites of a target polynucleotide, and (ii) result in a population of polynucleotides (e.g., a mutated population) in which many different sites of insertion or deletion are represented (i.e., different species in the mutated population comprise insertions or deletions at different sites). In contrast to the random mutations used in the present invention, a mutation is "directed" when it is made at a predetermined site in the polynucleotides in a population, such as the insertion of a cassette into a particular restriction site in the DNA segments of a population, or site-specific mutagenesis.

The art knows a variety of *in vitro* and *in vivo* methods for making random insertions and/or deletions in polynucleotides. Although it will be appreciated that the invention is not limited to any specific method for making insertions or deletions, illustrative examples of these methods are provided *infra*.

Usually the DNA segments to be mutated *in vitro* are closed circular molecules isolated from cells (e.g., plasmids, circular bacteriophage, and certain vectors) or, alternatively, may be circularized *in vitro*. Any method of circularization may be used. For example, linear bacteriophage, eukaryotic viruses, PCR products and other linear molecules can be circularized by treatment with DNA ligase or the equivalent. In some embodiments it will be desirable to carry out the ligation reaction at a low concentration of substrate molecules to avoid or reduce concatemerization. In certain embodiments, to limit nuclease activity to single cleavage event per molecule in the subsequent random linearization step (described *infra*) supercoiled circular DNA is used. Closed circular molecules can be supercoiled by treatment with topoisomerase II (Gellert et al., 1976, *Proc. Nat'l. Acad. Sci.* 73:3872-3876).

In one method of random mutation, the closed circular molecules are randomly cleaved, at a single site. A circular polynucleotide is "linearized" when it is cleaved once (in contrast to a polynucleotide that is "fragmented.") Methods for random linearization are known and include limited hydrolysis of double stranded DNA using double-strand cleaving nucleases (e.g., DNAse I) or using a combination of double-strand DNA nicking enzymes (e.g., DNAse I in the presence of ethidium bromide, topoisomerase mutants) and single-strand specific nucleases (e.g., S1 nuclease, P1 nuclease, Mung Bean nuclease). See, e.g., Yokochi et al., 1996,

Genes Cells 1:1069-1075; Chaudry et al., 1995, *Nucl. Acids Res.* 23:805-809. Alternatively, "pseudorandom" linearization can be carried out using a relatively non-specific restriction endonuclease (e.g., one that recognizes a common four base sequence) under conditions in which cleavage occurs approximately once per molecule. When necessary, prior to insertion or deletion, protruding ends may be blunted by filling in (e.g., using polymerase and dNTPs) and/or by treatment with exonuclease.

In practice, cleavage of a large population of molecules will usually result in a distribution of polynucleotides in addition to those that are linearized, including some molecules that are uncleaved, and others that are fragmented by cleavage at more than one site. It is known in the art to adjust enzyme and substrate concentrations, digestion times and other conditions to obtain primarily singly-cleaved molecules. If desired, linearized molecules can be isolated from fragments by routine methods (e.g., size selection by gel electrophoresis, chromatography, or centrifugation). However, it is not necessary to separate singly cleaved molecules from those that are uncleaved or multiply cleaved.

b) Random Insertions

The polynucleotide or oligonucleotide sequence(s) that are randomly inserted into a population of randomly linearized polynucleotides may be from any of a variety of sources. (The sequence(s) to be inserted can be referred to as the insertion sequence or the "insertion population.") Thus, the oligo/polynucleotides to be inserted may have a defined sequence(s) and/or biological function(s) (e.g., a *Drosophila* cuticle gene TATA box sequence). Polynucleotides suitable for insertion include defined functional modules or populations of modules (e.g., libraries of promoter, enhancer, or other regulatory elements, sequences encoding T- or B-cell epitopes, biotinylation domains, antibody selectable peptides, protein-binding domains, cellulose binding domains, selectable markers, reporter genes, protein loop sequences, functional domains of a protein, fragments of viral or bacterial genomes, and the like). Polynucleotides suitable for insertion also include defined or undefined fragments of molecules with a known function (e.g., fragments of a known promoter sequence, fragments of polypeptide coding sequences). The oligo/polynucleotides may be of unknown or random sequence and/or biological function, or may have no particular biological function in nature (e.g., a library of random sequence 12mers).

Suitable insertion polynucleotides may be generated by chemical synthesis, PCR

amplification, enzymatic fragmentation, or any other means. The size of the sequence(s) to be inserted may be in a wide range such as at least about 3, 6, 9, 12, 15, 18, 21, 25 or 50 bases in length up to about 0.1, 0.5, 1, or 2 kilobases or even larger. Insertion of the sequence between the termini of a linearized polynucleotide can be carried out by any suitable method. Typically the sequences to be joined are incubated together in the presence of a DNA ligase.

In some embodiments, a single species of polynucleotide (e.g., a 12-mer of a particular sequence) is randomly inserted into a population of polynucleotides. In different embodiments, a plurality (i.e., more than 1) of different species of polynucleotide is introduced in a particular step in the mutation process (e.g., a set of random sequence 12-mers, or a mixture of fragments of a promoter sequence is inserted).

The inserted sequences may modify or supplement the properties of the substrate molecules in any of a variety of ways. They may, as will be apparent from the examples provided *infra*, be selected to provide a particular sequence, such as a particular epitope coding sequence, protein binding or recognition site, transcription factor binding site, RNA splice site, or the like. Alternatively or in addition, they may act to introduce length variation into a polynucleotide or encoded polypeptide. In an encoded polypeptide, length variations influence the specificity of the molecule (e.g., substrate specificity in an enzyme, antigen specificity in an antibody). In a polynucleotide, length variation will, for example, change the spacing between transcription factor elements in a promoter, profoundly influencing the function of the promoter.

When insertions are made in a protein coding sequence of a polynucleotide, particular techniques can be utilized, if desired, to retain a particular reading frame (e.g., by insuring that the deletions and or insertions will be of a multiple of three nucleotide bases in length). For example, in one embodiment, a single codon (i.e., three nucleotides) is inserted. This can be accomplished by randomly inserting an oligonucleotide having a length that is a multiple of 3 bases (e.g., Boulain et al., 1986, *Mol. Gen. Genet.* 20:339-348). An alternative method involves first randomly inserting a resistance (e.g., drug resistance) cassette which can be cleaved out by restriction endonucleases after selection (e.g., growth on selective media). The insertion cassette can be designed to leave a single or multiple random or non-random codon(s) in the coding sequence (Wong et al., 1993, *Mol. Microbiol.* 10:283-292; Dykxhoorn et al., 1997, *Nuc. Acids Res.* 5:4209-4218; Hallet let et al., 1997, *Nuc. Acids Res.* 25:1866-1867). In addition, techniques for co-translational coupling of a reporter gene (e.g., GFP) may be used to identify or eliminate nonproductive (i.e., frame-shifted) products. It will be appreciated that

although retaining the original reading frame will reduce the number of "nonproductive" polynucleotides in the mutated population, and thus make screening somewhat more efficient, it is not necessary or always desirable to eliminate frameshift mutations.

c) Random Deletions

In some embodiments of the invention, deletions are introduced at random sites in a substrate population. The introduction of deletions may be used to reduce the size of a polynucleotide sequence (i.e., to increase the insert capacity of a vector), to change a property of a polynucleotide (e.g., by changing the spacing of functional domains in a polypeptide encoded by a DNA segment), and for other purposes.

When a population or polynucleotides is randomly deleted (i.e., deletions are introduced at random locations), there usually will be variation in the extent of deletions in various molecules in the population. The length(s) of deletions introduced in any one step will vary depending in the goals of the investigator, but will typically be less than 100 bases or basepairs (e.g., at least about 3, 6, 9, 12, 15, 18, 21, 25, 35, 50 or 100 bases in length). In some embodiments, however, some or all deletions may be longer, such as at least about 200 or 500 bases.

Deletions may be made by a variety of methods. In one embodiment, a circular or circularized molecule (e.g., a vector) is randomly linearized as described *supra*. The randomly linearized molecules are then reduced in size (i.e., sequence is deleted) by the use of a processive exonuclease (e.g., Bal31 or exonuclease III). In some embodiments, the resulting linear molecules are blunted by standard methods prior to recircularization by ligation (Sambrook et al., 1989, MOLECULAR CLONING - A LABORATORY MANUAL 2nd ed. Vol. 1-3). In one embodiment, sequences to be inserted (e.g., such as those described *supra*) can be included in the ligation reaction (resulting in simultaneous insertion and deletion of sequences relative to the substrate population).

In one embodiment of the invention, the polynucleotide is a vector and the introduction of random deletions and selection is used to reduce the size of the vector without eliminating sequences critical for the functioning of the vector (e.g., the replication origin). The reduced size increases the ability to introduce new or larger genes into the vector backbone. When using, for example, a bacteriophage vector with a limited DNA packaging length (due to capsid capacity), the reduction in size of the bacteriophage genome would allow the packaging

of new or larger genes without affecting essential phage functions. Notably, the present invention allows reduction in the size of a vector and/or introduction of genes from other sources without a priori knowledge of the function of parts of the parental vector. Thus, it is especially useful when using an uncharacterized bacteriophage as a vector (e.g., for use in *Streptomyces* bacteriophage ΦC31).

As noted *supra*, it will sometimes be desirable, when mutating a polynucleotide that encodes a polypeptide, to use techniques to retain a reading frame found in the parental vector. In one embodiment, for example, a single triplet is deleted from (each of) the deleted polynucleotides of a substrate population. This can be carried out by first inserting a resistance cassette which may be excised (e.g., after selection) deleting 3 nucleotides. For example, a cassette or short oligonucleotide containing a Type IIS restriction enzyme recognition site (e.g., EarI, SapI) can be designed which, after random insertion can be cleaved from the circular DNA so that a multiple of 3 nucleotides are removed. Alternatively, mobilization of a transposon (e.g., using cre/lox) may be used to excise the resistance cassette.

d) Additional Methods

In another embodiment of the invention, a mutated population is generated from a substrate population by the introduction of random insertion and/or deletions generated using processive exonuclease digestion of two subpopulations of polynucleotides. The subpopulations are then ligated to produce novel combinations of sequences, as described below.

According to this embodiment, the substrate population may be homogeneous (i.e., a plurality of polynucleotides having the same sequence, e.g., having the sequence of particular gene encoding a protein) or may be non-homogeneous (e.g., containing a mixture of polynucleotides having related sequences, such as a family of related genes [e.g., encoding human actins] or homologs from different species [e.g., encoding human and bovine actin genes], or the product of shuffling reactions, or other non-identical polynucleotides as described supra).

To produce a mutated population having random insertions and/or deletions, the substrate population is divided into at least two subpopulations. A series of nested deletions is produced from each of the, e.g., two subpopulations by incubation with exonuclease using methods well known in the art (see, e.g., Henikoff, 1984, *Gene* 28:351, see also New England Biolabs Catalog 1998/99 page 129 "Exo-SizeTM Deletion Kit"). Briefly, a nuclease such as exonuclease III is used to create unidirectional deletions in the polynucleotides of each

subpopulation. Preferably, restriction endonuclease digestion of the DNA segments in each subpopulation is used to introduce both a nuclease susceptible end (i.e., a 5' overhang or blunt end) and a nuclease nonsusceptible end (i.e., a 3' overhang) such that the nuclease digests in only one direction. The at least two subpopulations differ in that the site of the nuclease susceptible end is different in different subpopulations. After a series of deletions of varying lengths (i.e., nested deletions) is produced in each subpopulation (e.g., by incubating aliquots with exonuclease for differing lengths of time) polynucleotides from each subpopulation are ligated to produce a mixture of mutated polynucleotides having random insertions (e.g., duplications) and/or deletions at the junction site (a mutated population).

An example will help to illustrate this embodiment of the invention. Thus, consider a homogeneous substrate population of DNA segments encoding a polypeptide, which substrate population is divided into two subpopulations. In one embodiment of the method, the nuclease susceptible end in one subpopulation is introduced at the polynucleotide site corresponding to the amino-terminus of the encoded polypeptide with digestion toward the c-terminus, and the nuclease susceptible end in the other subpopulation is introduced at the polynucleotide site corresponding to the carboxy-terminus of the encoded polypeptide, with digestion toward the n-terminus. For purposes of description, the two subpopulations in this illustrative example can be referred to as producing a "amino-terminus deleted" product or a "carboxy-terminus deleted" product.

After a series of nested deletions is produced in each subpopulation, polynucleotides from each subpopulation are ligated to produce a mixture of mutated polynucleotides having random insertions (e.g., duplications) and/or deletions at the junction site. Thus, continuing with the example provided above, and by way of illustration, and not limitation, imagine that in each of the subpopulations deletions range from 1 base to about 99% of the length of the polynucleotide (including, e.g., 5%, 10%, 90% and 95% deletions). It will be appreciated that the ligation of an amino-terminus deleted molecule from which exactly 10% of the length of the molecule is deleted to a carboxy-terminus deleted molecule from which exactly 95% of the length of the molecule is deleted will result in a molecule that has a 5% duplication (at the ligation junction) compared to the substrate polynucleotide sequence. Likewise, the ligation of a amino-terminus deleted molecule from which exactly 5% of the length of the molecule is deleted to a carboxy-terminus deleted molecule from which exactly 90% of the length of the molecule is deleted will result in a molecule that has a 5% deletion (at the ligation

junction) compared to the substrate polynucleotide sequence.

It will be apparent that many variations of this basic scheme are available, including, for example, introduction of susceptible ends at sites other than those corresponding to polypeptide termini.

It will be appreciated that the present invention is not limited to any particular method of random insertion or deletion, and that methods other than those specifically described *supra* may be used. For example, self inserting DNA, i.e., transposons, may be used for *in vivo* insertion combined with a subsequent *in vivo* excision by mobilization, or *in vitro* excision by restriction endonucleases.

It will often be desirable, prior to the screening step (*infra*), to enrich the mutated population(s) for polynucleotides that have been mutated (i.e., by insertion or deletion). Enrichment is desirable because even efficient methods for insertion and deletion will often result in a mutated population containing some molecules, or even a substantial proportion of molecules, that are wild-type (i.e., do not contain an insertion or deletion). Using an enrichment step will reduce the size of the population that must be subsequently screened. A variety of methods can be used for enrichment. One method, the use of resistance cassettes, is discussed *supra*. Another suitable method for enrichment of insertion events is carried out by denaturing the DNA of the mutated pool, and subsequently binding it to another aliquot of the inserted DNA which is immobilized on a solid support. Unbound (e.g., wild-type) polynucleotides are removed by washing and the mutated molecules are eluted from the affinity matrix (e.g., using temperature, urea, etc.). Another suitable method for enrichment involves inserting an oligo- or polynucleotide that contains, in addition to the sequence to be inserted, a second sequence, such as a *lac* operator site, that is bound by an immobilized sequence specific DNA-binding protein (e.g., the LacI repressor). After washing, polynucleotides with the insertion can be eluted (e.g., in the presence of isopropylthiogalactoside). Subsequently the oligo- or polynucleotide sequence responsible for binding can be excised from the polynucleotide, if desired, by a variety of methods, (some of which are discussed *supra)*, leaving behind the sequence to be inserted.

It will be apparent from the description *supra* that the practice of the invention involves various techniques well known to persons of skill in the art of molecular biology. Instructions sufficient to direct persons of skill through appropriate cloning, sequencing, mutation, random recombination techniques, and other techniques found in, e.g., Berger and

Kimmel, Guide to Molecular Cloning Techniques, METHODS IN ENZYMOLOGY volume 152 Academic Press, Inc., San Diego, CA; Sambrook et al. (1989) MOLECULAR CLONING - A LABORATORY MANUAL (2nd ed.) Vol. 1-3; and CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, F.M. Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (1998 Supplement), and other references cited herein and other references known in the art.

IV. Screening a Mutated Population

Another step in the method of the present invention is the screening of a mutated population for a desired property. This results in the identification and isolation of, or enrichment for, DNA segments that acquire the desired property as a result of the mutation (e.g., a *new* property), or in which an existing property is desirably enhanced. As used herein, the term "screening" has its usual meaning in the art and is, in general, a two-step process. In the first step it is determined whether a DNA segment has a particular property and in the second step the DNA segment(s) with the property are physically separated from those not having the property. For convenience, the population of polynucleotides resulting from the screen may be referred to as the "selected population."

In some forms of screening, identification and physical separation are achieved simultaneously. For example, identification and separation of a polynucleotide conferring drug resistance to a cell can be accomplished by selection of cells resistant to the drug (e.g., culturing under conditions in which non-resistant cells do not survive). It will be clear from this example, that the "separation" step of screening does not imply or require isolation of a biochemically pure polynucleotide with the desired property. Rather, separation means that the DNA segment of interest is separated from other DNA segments (e.g., cells comprising other DNA segments). In some embodiments of the invention, when screening is carried out, the physical separation of DNA segments with the property and those without need not be absolute, and due to methodological limitations often is not. Thus, in some embodiments, the screening of the mutated population results in a selected population that is enriched for the DNA segments with the desired property.

It will be immediately apparent to those of skill that screening requires an assay to identify DNA segments having the desired property. It will also be apparent that the specific assay will depend upon the particular desired property. A variety of examples are provided *infra*

to provide additional guidance to those of skill. Numerous additional screens suitable for use in the present invention are described in publications and disclosures describing "DNA Shuffling" methods. Thus, the reader is referred to the patents, applications, and publications listed in the Section I, *supra*, in the description of "shuffling," each of which is incorporated herein by reference in their entirety and for all purposes. It will be appreciated, however, the invention is not limited to any particular screening method.

V. Recursive Mutation and Screening

In one embodiment of the invention, the selected population, generated as described *supra*, is mutated, i.e., insertions, deletions or both are introduced at random sites in the DNA segments in the selected population. The type of mutation may be the same or different from the mutations introduced into the substrate population (i.e., the original or first substrate population). For example, in a case in which random insertions were made in the substrate population, insertions may also be introduced in the selected population or, alternatively, deletions may be introduced. Moreover, when insertions are made, the polynucleotide inserted may be the same or different from the insertion polynucleotide in the previous step. The resulting population of mutated DNA segments may be referred to as a "recursively mutated population" in reference to the fact that the DNA segments have been subjected to more than one cycle of mutation by insertion and/or deletion.

The recursively mutated population is then screened for the desired property. The population of DNA segments resulting from this screen is referred to a "recursively selected population" (i.e., a "first recursively selected population"). The screen used for the "selected population" and the "recursively selected population" may be the same or different. In embodiments in which the same screen is used, the stringency of the screen will be increased to identify DNA segments with increasingly robust properties. For example, if the desired property is the ability (of a DNA segment) to confer drug resistance to a cell, the second or subsequent screening assay may use a higher concentration of the drug than the initial screen (i.e., the screen of the mutated population). As another example, if the desired property is the ability of a DNA segment to encode a polypeptide that is bound by a particular antibody, increasingly stringent binding conditions may be employed in screens.

As illustrated in Fig. 1, additional cycles of mutation and screening may be carried out, if desired. Generally, from 1 to 50 additional cycles will be carried out, more often from

about 3 to about 10 additional rounds. In cases in which additional cycles of mutation and screening are carried out, it is convenient to refer to the resulting selected populations as the "second recursively selected population," the "third recursively selected population," etc.

As is evident, each of the recursively selected populations contain DNA segments with the desired property. Although in some cases the population as a whole will be useful, more often a particular species of DNA segment will be isolated from the population and used.

VI. Mutation of Multiple Substrate Populations and Screening of Recombinants

In a related embodiment of the invention, random insertions or deletions are introduced into two (or more) different substrate populations and sequence elements from each population are combined by directed recombination or random recombination (e.g., shuffling). Typically, different insertion sequences are introduced into each of the substrate populations. One or each of the mutated substrate populations may be subjected to screening or selection for a particular property conferred by the mutation of that population, prior to the recombination of the substrate populations. Whether or not screening of the mutated substrate populations is undertaken, the recombined population will be subjected to screening/selection for the desired property or combination of properties.

As noted, random recombination methods include DNA shuffling techniques. Shuffling can be carried out in conjunction with the introduction of point mutations (e.g., by error-prone amplification), or without introduction of point mutations (e.g., by the use of proofreading polymerases). In contrast, "directed recombination," or subcloning, refers to methods of recombination that require knowledge of the restriction map of at least part of each substrate population and result in the insertion of a restriction fragment from one population in to a particular restriction site in the second population. Examples include the insertion of particular restriction fragments (by restriction and ligation) or PCR amplicons (usually by ligation or SOE-PCR ["splicing by overlap extension- PCR"]) derived from one substrate population into a specific site or location in the second substrate population, and ligation of two randomly linearized substrate populations.

VII. Random Recombination of the Selected Population

In a different embodiment of the invention, the selected population (described in §III, *supra*), a recursively selected population (described in §V), or a DNA segment species

isolated from such a population is used as the starting material for methods which lead to random recombination and point mutation, e.g., DNA shuffling. It will be understood that *random* recombination refers to recombination methods other than directed exchange of specific defined sequences (e.g., the transfer of a sequence from one population of DNA segments to a second population by restriction and ligation of defined restriction fragments, for example as described in Section VI, *supra*). Random recombination methods rely instead on the generation of a large pool of DNA fragments by random fragmentation of a single DNA sequence or a family of related DNA sequences, and the reassembly of the fragments in various combinations to produce DNA segments with a new structure (i.e., new combinations of deletions, insertions and/or introduced point mutations) and with the desired property.

Recursive random recombination or non-recursive random recombination methods may be used. The term "recursive" in this context refers to the use of multiple cycles of fragmentation, recombination, and screening (e.g., at least 2, sometimes at least 5 cycles). Typically, when a random recombination method is applied to a single DNA segment from a selected population, a recursive recombination method will be used, e.g., Zhang et al., 1997, *Proc. Natl Acad. Sci.* 94:4504. When a population of different DNA segments are used, both recursive and non-recursive recombination methods (i.e., a single cycle of fragmentation, recombination, and screening) are suitable (see, Crameri et al., 1998, *Nature* 391:288-291).

VIII. Exemplary Applications

This section provides several exemplary examples to illustrate various uses of the invention. Numerous other uses and variations will be apparent to one of skill upon reading the present disclosure.

Exemplary Application 1: Changing Promoter Specificity

In one embodiment, the methods of the invention are used to evolve a transcription regulatory sequence (e.g., a promoter or enhancer sequence) so that the expression characteristics of the regulatory sequence, such as inducibility, tissue specificity, or promoter strength are changed. The use of the methods of the invention is particularly powerful for the evolution of regulatory elements, because such elements are typically modular in structure, with different combinations of modules (or differences in relative orientation) contributing to regulatory activity/function in unpredictable ways.

Typically the mutation and screening of a promoter sequence is carried out using a vector (e.g., an expression vector) in which the target promoter is operably linked to a reporter gene (i.e., a gene encoding a gene product that can be conveniently assayed). Many suitable reporter genes are well known in the art, including the green fluorescent protein (GFP), luciferase, β-glucuronidase, β-galactosidase, and secreted alkaline phosphatase. An advantage of using a promoter-reporter system is that a change in promoter function can be easily detected, facilitating a variety of simple screening methods. Once the promoter sequence is evolved by the present method to have the desired property or combination of properties, the promoter region can be cloned into a different vector (e.g., to drive transcription of a gene of interest other than the reporter gene). Alternatively, the reporter-gene sequence can be removed from the mutated vector and a different gene of interest inserted in its place. Methods for subcloning a promoter or coding sequence in a vector are well known to those of skill in the art (see, e.g., Ausubel et al., *supra*). For example, the mutated promoter can be amplified by the polymerase chain reaction and the amplified sequence cloned into a region upstream of a selected coding sequence.

Thus, in one exemplary embodiment of the invention, (1) the substrate population is a population of DNA segments having a particular promoter activity (e.g., the ability to direct transcription of a reporter gene in a hepatocyte specific manner) and (2) the desired property is a different promoter activity (e.g., the ability to drive expression in T lymphocytes) or combination of activities (e.g., the ability to drive expression in both T lymphocytes and hepatocytes, but not pancreatic beta-cells). The generation of a lymphocyte-specific promoter, for example, may be carried out by mutating a substrate population comprising a hepatocyte promoter operably linked to a GFP reporter gene, and carrying out a suitable screen of the resulting mutated population.

The promoter sequences are mutated by random insertion and/or random deletion. As described *supra*, examples of suitable polynucleotides for insertion include random fragments from known promoters (e.g., a T-cell or hepatocyte specific promoter, the metallothionein promoter, the constitutive adenovirus major late promoter, the dexamethasone-inducible MMTV promoter, the SV40 promoter, the MRP polIII promoter, the constitutive MPSV promoter, the constitutive CMV promoter, and promoter-enhancer combinations known in the art), synthetic oligonucleotides constituting modules from known promoters, random sequence polynucleotides, and other sequences. In embodiments in which there is more than one round of mutation, different polynucleotides may be inserted at different steps. For example, the substrate

population may be mutated by random insertion of random fragments of a MMTV promoter element and the selected population may be mutated by random insertion of a defined fragment from a metallothionein promoter.

One suitable screen comprises transducing the mutated population of polynucleotides into cultured cells of a particular type (e.g., a Jurkat T lymphocyte cell line), assaying reporter gene expression in the cells (for example by using fluorescence activated cell sorting to detect GFP expression), and selecting cells in which the reporter gene is expressed. Expression in the Jurkat cell type indicates that the mutated hepatocyte promoter segment has acquired the ability to drive transcription in the second cell type. The mutated DNA segments may then be isolated from the population of transduced cells showing the desired property (e.g., new expression specificity), pooled (if not isolated as a pool), and used for additional round(s) of random insertion/deletion mutagenesis or random recombination. Subsequent rounds of mutation and screening may be used to evolve a subpopulation with a higher GFP expression level in Jurkat cells, to add other elements to the promoter (e.g., conferring steroid hormone inducibility). Additional screens may be carried out, if desired, to identify novel promoters with additional desired characteristics. For example, following or concurrently with a screen for the ability of the mutated DNA segments described above to drive expression in T cells, it may be desired to transduce the DNA segment population into hepatocytes and screen for the ability (or lack of ability) to drive transcription in hepatocytes. Using combinations of screens, it is possible to identify novel promoter sequences that, for example, drive expression in T cells and hepatocytes, but not beta-cells. Additional panels of cells types and other variations will be evident to one of skill upon reading this disclosure.

It will be recognized that in the screens described above, control experiments, which will be known to those of skill, will usually also be carried out. If desired, the DNA segment having the new transcription specificity can be isolated from the cell for further manipulation (e.g., it can be operably linked to a variety of coding sequences).

As will be apparent to those of skill, when the mutation step is carried out on a vector comprising the promoter and reporter gene, some of the mutations may disable the reporter gene function (e.g., by introducing a frame shift). In such a case, the "non-productive mutants" in the mutated population will be eliminated in the screening step. Alternatively, the mutation steps may be carried out on a vector containing the promoter only, and following mutation the promoter sequences can be transferred (e.g., by restriction and ligation and/or PCR

amplification of the promoter sequence and insertion of the product) as a cassette into a pristine vector comprising a reporter gene. A variety of strategies will be apparent to one of skill following the guidance of this disclosure.

Exemplary Application 2: Changing an Enzymatic Activity

In some embodiments of the invention, the substrate population is a population of DNA segments encoding a polypeptide with an enzymatic activity and the desired property is a new enzymatic activity. In one embodiment, the substrate DNA segments encode a polypeptide with β-galactosidase activity, and the different enzyme specificity desired is fucosidase activity. Recursive rounds of mutation by alternative deletions (of 5-20 basepairs) and insertions (from a library of random hexamers) can be combined with a screen as described in Zhang et al., 1997, *Proc. Nat'l Acad. Sci.* 94:4504. As noted *supra*, in cases in which protein coding DNAs are mutated it will often be desirable to use mutation methods that retain the existing reading frame (e.g., deletion and/or insertion of a multiple of 3 nucleotide bases), although, if desired, non-functional frame-shift mutants can be eliminated during the screening step.

Exemplary Application 3: Changing a Property of an Encoded RNA

The methods of the invention may be used to evolve a regulatory element (or other region) of an RNA encoded by the DNA segment. For example, RNA stability elements are known which confer increased stability on mRNAs with which they are physically associated (e.g., encoded downstream of the protein coding sequence). Thus, in one embodiment of the invention, the substrate population is a population of DNA segments that encode mRNA, and the desired property is increased mRNA stability.

The evolution of a mRNA-encoding sequence to encode a more stable RNA is accomplished by randomly inserting DNA sequences into a substrate population encoding an mRNA, and screening or selecting for high levels of expression of the protein (because, generally, expression of the protein product of the gene is proportional to the mRNA stability) or directly assaying the expression level of the mRNA. In one embodiment, the inserted sequences are fragments (e.g., defined or random fragments) of DNA sequences from known stability elements (Chan et al., 1998, *Proc. Nat'l Acad. Sci.* 95:643-6547; Russell et al., 1998, *Mol. Cell. Biol.* 18:2173-2183).

In one embodiment, the increased gene expression in the mutated population is detected and the resulting set of clones (or pools of 2-20 clones having the highest mRNA stability), i.e., the selected population, is used in shuffling or, as a target population for additional mutation. The additional mutation can include insertion of additional downstream mRNA stability conferring fragments (the same as or different from those inserted in earlier steps), deletion and screening for increased mRNA stability, or the insertion of different sequences (e.g., to confer a different selectable property on the RNA-encoding DNA segment).

Exemplary Application 4: Addition of a Functional Domain to a Cloning or Expression Vector

In this example, the DNA segments of the substrate population are cloning vectors which may be procaryotic, eukaryotic, or shuttle vectors, and which may be characterized vectors (e.g., pUC18) or uncharacterized vectors. Examples of vectors include artificial chromosomes, plasmids, episomes, viruses, bacteriophages, and mobile elements (e.g. transposons, insertional elements). It is often desirable to add a new functional domain or element to a vector by inserting a cassette encoding a polypeptide (e.g., encoding a resistance marker or novel gene of interest), regulatory element, combinations of genes and regulatory elements, or other functional or structural elements. However, often the optimal location for insertion is not known. It is especially difficult to design vectors with particular or optimal properties when the vectors are complex (e.g., human papilloma virus and other eukaryotic viruses) or intended for use in relatively uncharacterized species of fungi, plants, bacteria (e.g. *Streptomycetes*), etc.. By inserting the function domain, or a fragment thereof, in a random manner, screening the resultant mutant population and optimizing the desired property(s) by recursive insertion/deletion mutation (and, optionally, shuffling), it is possible to efficiently generate vectors with novel and optimized properties.

In one embodiment, an expression cassette (e.g. GFP under control of the *E. coli lac* promoter) is inserted into random positions of the pool of a mixture of randomly linearized vectors (e.g., a pool of pUC19, pET11, pBR322, and pBAD24). Following transformation into host cells (e.g., *E. coli*) the expression of the protein is assayed (e.g., as assessed by its activity, e.g., green fluorescence for GFP), and the clones expressing the highest levels of the reporter gene when induced by IPTG or arabinose are identified and isolated (see, e.g., Crameri et al.,

1996, *Nature Biotech.* 14:315-319). DNA shuffling and further screening is carried out. The resulting product is a vector comprising the GFP structural gene positioned in a particular vector backbone at a position that provides the best expression properties of the protein.

Exemplary Application 5: Building an Operon Conferring a Multigenic Phenotype on Cells

In another example, the methods of the invention are used to generate a bacterial operon encoding several coding sequences (e.g., genes encoding proteins active in a particular metabolic pathway). Thus, in one embodiment, the coding sequences for each of the polypeptides (e.g., enzymes) to be expressed is inserted in a stepwise fashion (e.g., as outlined in Figure 1) into a vector comprising one or more promoters able to drive transcription of the polypeptide coding sequences. After each insertion step, a screen is carried out for cells optimally expressing the phenotype conferred by the inserted polypeptide(s). The resulting multigenic operon comprises each of the polypeptide sequences positioned relative to each other, regulatory elements, and other vector elements in positions that result in optimal expression (or other selected-for properties).

Exemplary Application 6: Insertion of an Affinity Selectable Tag into a Polypeptide

In another example, a cassette encoding an affinity selectable tag is randomly inserted into a substrate population of DNA segments that comprise a polypeptide coding sequence, resulting in mutant polypeptides that retain biological activity and have acquired the ability to be affinity selected. The addition of an affinity selectable tag to a biologically active protein is useful for, e.g., protein purification.

Examples of sequences that can be randomly inserted into the polypeptide coding sequence of the substrate population include polynucleotides encoding affinity selectable oligo- or polypeptide sequences (e.g., peptide epitopes recognized by an immunoglobulin), anti-antibody fragments (e.g., Vaughan et al., 1996, *Nat. Biotech.* 14:309-314) and others well known in the art. Following insertion, the mutated population is screened and/or selected by a combination assays: typically one assay identifies mutant polypeptides that include the affinity selectable sequence and a second assay identifies polypeptides that have a second biological property (such as the ability to encode a catalytically active enzyme). Screening for affinity (affinity selection) may be carried out by any suitable method, such as affinity chromatography,

immunoprecipitation, etc. In some embodiments, a phage display system is used for affinity enrichment. In such systems, the encoded oligo- or polypeptide is presented on the surface of a cell, virus or bacteriophage where it is susceptible to binding by the affinity partner (see e.g., Ernst et al., 1998, *Nucleic Acids Res.* 26:1718-1723; and U.S. Patent Nos. 5,223,409 and 5,403,484).

Exemplary Application 7: Production of Protein Vaccines

The production of protein vaccines is very often limited by the inefficient expression of the antigenic protein or inefficient processing of the antigen for presentation on MHC complexes. This can be overcome by insertion of one or several epitope sequences from the antigen into a well expressed or efficiently processed protein. Thus, in one approach, multiple T-cell and/or B-cell epitopes are inserted into a known protein "scaffold." In one embodiment, the present invention is used to produce effective vaccines by the insertion of immunodominant T-cell and B-cell epitopes of an immunogenic protein in the scaffold of a highly expressible protein.

In an exemplary embodiment, a known B-cell epitope from HIV gp120 is inserted into a human scFv protein (Vaughan et al., 1996 *Nature Biotechnology* 14:309-314) and expressed in *E. coli*. The presence of the B-cell epitope in the chimeric protein is screened for as described in copending USSN 09/021769 and 60/074,294. Positive clones (i.e., from the selected population) are pooled and all positive clones are used for the next round of insertion of additional B-cell epitopes and/or T-cell epitopes. DNA shuffling is carried out using DNA from individual clones. The resulting polypeptide comprises multiple well-expressed and well-processed immunogenic peptides and is useful as a vaccine.

IX. EXAMPLES

The following examples are provided to illustrate the practice of the invention.

EXAMPLE I

Synthesis of a Bacterial Vector Containing a New Regulatable Promoter

This example demonstrates the use of the invention to produce a vector with novel properties. Beginning with a known vector (pAK400-GFP) capable of expressing green fluorescent protein (GFP), a process including two cycles of random insertion/deletion

mutation and selection or screening are used to produce a panel of novel vectors. The new vectors have new (compared to the parental vector) desired properties with respect to tetracycline resistance, inducibility, and GFP expression levels.

A) Synthesis of Randomly Linearized pAK400-GFP

The parental vector pAK400-GFP is based on the pAK400 vector (Krebber et al., 1997, *J. Immunol. Meth.* 201:35-55), but is modified by replacement of sequences encoding the tet^R (tetracycline resistance) gene with the coding sequence for green fluorescent protein (GFP). To construct pAK400-GFP, GFP is PCR amplified by primers "GFP.For" and GFP.Rev" from pBADGFP cycle 3 (Crameri et al., 1996, *Nature Biotech.* 14:315-319) and cloned by *Nde*I and *Hind*III in a three fragment ligation into a *Nde*I and *Hind*III vector fragment of pAK400, resulting in "pAK400-GFP." In pAK400-GFP, expression of GFP is under the control of the *lac* promoter and is inducible by isopropylthiogalactoside (IPTG). The vector also contains an *E. coli* pUC derived ColE1 origin of replication, a *lac*I gene for the expression of the lac repressor in order to repress the lac promoter efficiently, an f1 origin for packaging of single stranded DNA in phagemids, and the gene for chloramphenicol acetyl transferase which confers resistance to chloramphenicol (Cam^R).

Supercoiled pAK400-GFP is prepared in *E. coli* by CsCl/ethidium bromide equilibrium centrifugation according to standard procedures (e.g., Sambrook et al., *supra*). The vector is linearized by random cleavage by treatment with DNAse I in the presence of ethidium bromide, as described in Chaudry et al., 1995, *Nucleic Acids. Res.* 23:3805-3809. Following phenol/chloroform extraction, the once randomly nicked vector is treated with S1 nuclease at low pH to cleave opposite the single stranded nick (Chaudry et al., *supra*). The randomly linearized vector is extracted using phenol/chloroform, precipitated and treated with a polymerase (to ensure the DNA is blunt ended) and with alkaline phosphatase (to dephosphorylate the linearized molecules to prevent self-ligation). Finally the linearized (i.e., once cleaved) molecule is purified on a 5% polyacrylamide gel or by CsCl/ethidium bromide equilibrium centrifugation (Sambrook et al., *supra*).

B) Synthesis of tetR polynucleotides for random insertion

The *tet*RA operon containing the tet^R (tetracycline resistance) gene of Tn10 (Schollmeier et al., 1984, *J. Bacteriol.* 160:499-503) is PCR amplified from pAK400 (Krebber

et al, 1997, *J. Immunol. Meth.* 201:35-55) using the phosphorylated primers Tet.For and Tet.Rev and a proof-reading polymerase (*Pfu*; Stratagene).

C) Inserting randomly the tet operon into pAK400-GFP

The blunt ended products of (A) and (B), *supra*, are ligated to each other according to standard procedures (Sambrook et al., *supra*).

D) Selecting for tetracycline and chloramphenicol resistance and screening for inducibility of GFP by IPTG

The ligation reaction of step (C) is transformed into an *E. coli* K12 strain. The transformed cells are plated and selected on LB agar containing chloramphenicol, tetracycline and IPTG ("IPTG plates"). After growth overnight at 37°C, colonies are selected on the basis of green fluorescence upon exposure to UV light (Crameri et al., 1996, *Nature Biotech.* 14:315-319), indicating expression of GFP. The GFP-expressing colonies are replica plated onto agar plates containing chloramphenicol, tetracycline, and 2% glucose ("glucose plates") and assayed for GFP expression (by inspection under UV irradiation). DNA is prepared from 100 colonies that express GFP on IPTG plates (initial plating) but not on glucose plates (replica plating). These DNA segments compromise a population of different (in respect to the position of the *tet*RA-operon) vectors with the phenotype: CamR, Tet^R, IPTG-inducible expression of GFP (i.e., IPTG inducible promoter). The vectors in this population may be referred to as pAK400-GFP-Tet. As noted *supra*, the tetR gene is inserted in different positions in different species in the population.

E) Synthesis of double stranded oligonucleotides from the tet regulatory unit of Tn10

Non-phosphorylated double-stranded oligonucleotides (the pairs of Op1.For/Op1.Rev and Op2.For/Op2.Rev) which encode the two operators of the tn10 promoter (Bertrand et al, 1983, *Gene* 23:149-156) are synthesized chemically. Together the two oligonucleotides are referred to as the "tet oligonucleotides."

F) Ligation of the tet oligonucleotides into the linearized vector pAK400-GFP and swapping of the promoter region into pAK400-GFP-Tet

In this and the following steps, the tet oligonucleotides are randomly inserted

into linearized pAK400 vector (linearized as described for the pAK400-GFP vector in step A, *supra*, but not dephosphorylated) to produce a population of pAK400 vectors containing random insertions of the oligonucleotides. Subsequently the (mutated) lac promoter regions from the population (containing insertions) are transferred to the population of pAK400-GFP-Tet vectors made in step D, *supra*.

(An alternative strategy would be to randomly insert into the pAK400-GFP-Tet vector population. The strategy used is preferred because it requires screening fewer clones, i.e., only clones in which the tet oligonucleotides have inserted at random sites within the *lac* promoter region rather than in other sites in the vector.)

As a first step, the concentration of double stranded tet oligonucleotides is optimized by ligating different amounts of oligonucleotide into the randomly linearized vector, followed by transformation into an appropriate *E. coli* K12 strain. After growth overnight at 37°C, the colonies are counted. The optimal concentration of oligonucleotide is that concentration that just decreases the number of colonies. Although optimizing the oligonucleotide concentration will increase efficiency, this step is not critical.

Having determined the optimal oligonucleotide concentrations for insertion into the randomly linearized pAK400 (from above), the double-stranded tet oligonucleotides encoding parts of the tet promoter region are inserted into the randomly linearized pAK400 vector by blunt end ligation. After phenol/chloroform extraction, the resulting ligation is cut with *Kpn*I and *Nde*I at unique sites flanking the lac promoter of pAK400. The resulting fragments containing the *lac* promoter and a *tet* promoter oligonucleotide are isolated using electrophoresis in a non-denaturing 8% polyacrylamide gel (Sambrook et al., *supra*). The *Kpn*I-*Nde*I fragment from pAK400 is 209 bp. When a 20 basepair oligonucleotide is inserted, the *lac* promoter fragment will increase in size to 229 bp. Accordingly, a 229 bp band is isolated from the non-denaturing gel. The isolated fragment is cloned (ligated) into the pAK400-GFP-TET vector pool, which has been *Kpn*I and *Nde*I digested. The result is that some (though usually not all) of the resulting ligation products will comprise a randomly mutated *lac* promoter (i.e., containing random insertions of the tet promoter oligonucleotide) in a pAK400-GFP vector that is also randomly mutated (i.e., by random insertion of *tet*RA operon).

G) Selecting for tet and cam resistance and screening for inducibility of GFP by IPTG and/or tetracycline

The ligation of step (F) is transformed into an appropriate *E. coli* K12 strain. The transformation is plated and selected on agar plates containing 30μg/ml chloramphenicol, 5μg/ml tetracycline, and 2% glucose. The colonies are grown overnight at 37°C.

The recombinants are screened to identify vectors which have different promoters. The expression of GFP in the presence and absence of IPTG and/or tetracycline is determined as described *infra*. Tetracycline and chloramphenicol resistant colonies are selected by growth in the presence of these two antibiotics. The resistant colonies are replica plated on to four different plates. All plates contain chloramphenicol (to select for the Cam^R of the pAK400 vector backbone). Plate 2 additionally contains IPTG, Plate 3 additionally contains tetracycline, and Plate 4 additionally contains tetracycline and IPTG.

Expression of the GFP reporter gene by colonies is detected by visual or electronic observation of green fluorescence of colonies exposed to UV light (Crameri et al., 1996, Nature Biotech. 14:315-319). Colonies that express GFP on one plate and not on one of the others are regulated by either IPTG and/or tetracycline. Compared to the parental vector (which is exclusively regulated by the presence or absence of IPTG) colonies in which GFP expression is either increased or decreased by the presence or absence of tetracycline have a regulatory function not present in the parent. This screen is able to identify populations of vectors with new phenotypes, i.e., Cam^R, Tet^R, and GFP expression when different combinations of tetracycline and IPTG are used.

The described properties of these vectors may be enhanced further by additional rounds of insertion, rounds of deletion, or by shuffling, using the same screen described *supra* (and, e.g., assaying for increased levels of GFP expression) or other screens.

EXAMPLE II

Production of a β-Lactamase Containing an *In Vivo* Biotinylation Peptide

This example demonstrates the generation of a high-activity beta-lactamase polypeptide that contains an *in vivo* biotinylation sequence. The beta-lactamase gene is capable of conferring ampicillin resistance when expressed in a bacterium; the biotinylation sequence may be used to detect or purify a polynucleotide comprising the high-activity beta-lactamase polypeptide. This example is illustrative of the creation of a novel multifunctional polypeptide

using the techniques of the invention.

A) The *bla* gene (encoding beta-lactamase) is PCR amplified from pUC19 using the primers Bla.For and Bla.Rev and subsequently cloned into the SfiI restriction site of pAK200 (Krebber et al., 1997, *J. Immunol. Meth.* 201:35-55). The resulting vector, pAK200SAMP is randomly linearized (but not phosphorylated) as described in Example I, *supra*.

A double-stranded 90-bp polydeoxyribonucleotide is generated by annealing of 90-mers Bio.Rev and Bio. For (encoding a polypeptide having an *in vivo* biotinylation site sequence (Schatz, 1993, *Bio/Technology* 11:1138-1143), added in excess, and ligated to the randomly linearized pAK200SAMP vector at random positions. The *in vivo* biotinylation site becomes biotinylated when the protein is expressed in *E. coli* strains which express the endogenous biotin holoenzyme synthetase encoded by *bir*A (Barker et al., 1981, *J. Mol. Biol.* 146:451-467).

The pAK200SAMP vector is cleaved with *Sfi*I. The fragment containing the bla gene and a 90 bp insertion is identified by size and gel purified by standard methods. The fragment including the biotinylation sequence is approximately 896 bp (compared to approximately 806 bp without the insert). The purified fragments are cloned into the *Sfi*I site of phage display vector pAK200 (Krebber et al., 1997, *supra*). After transformation of the phagemid library, the bacteria are spread on 2YT-agar plates containing 30μg/ml chloramphenicol and a concentration of ampicillin that reduces the recovery from the transformation to 50% of the measured complexity (measured complexity is assessed by plating on 2YT-agar containing 30μg/ml chloramphenicol; hereinafter "2YT-Cam30" plates).

After growth overnight at 30°C, the plates are scraped and resuspended in 2YT. An aliquot is added to 100 ml 2YT-Cam30 containing the above calculated concentration of ampicillin. After coinfection with VCSM13 (Stratagene) according to Krebber et al., 1997, *supra,* and growth, the phages are precipitated and panned in PBS/dialyzed 2% skim milk for two to four rounds against streptavidin (Hawkins et al., 1992, *J. Mol. Biol.* 226:889-896) immobilized on magnetic beads (Dynal). The binding of single clones to streptavidin is verified by phage ELISA (Lindner et al., 1997, *Biotechniques* 22:140-49). These clones (which are heterogeneous) are referred to as "pAK200-bla-bio." The combination of the selection on ampicillin plates and the panning procedure identifies polynucleotides encoding an active beta-

lactamase gene containing a biotinylation sequence.

B) The expression and beta-lactamase activity of the pAK200-bla-bio produced in Section A, *supra*, is optimized by PCR shuffling (Stemmer, 1994, *Nature* 370:389-391). To do this, five to ten pAK200-bla-bio species (clones) are selected based on comparatively high beta-lactamase activity (as assessed by conferring on host bacteria resistance to high ampicillin concentrations). The bla-bio insertion is amplified by PCR using Bla.For and Bla.Rev primers. According to a standard PCR shuffling protocol (Stemmer, 1994, *Nature*, *supra*), the PCR products are fragmented randomly by DNAse I, reassembled and cloned into the *Sfi*I sites of pAK200SAMP. The library is grown overnight at 30°C on 2YT-Agar containing 30μg/ml chloramphenicol and a concentration of ampicillin (the "limiting" concentration) which reduces the recovery from the transformation to 25% of the measured complexity when grown on plates lacking ampicillin. As described *supra*, the library is scraped from the plates, grown in the presence of the limiting concentration of ampicillin, and coinfected with helper phage (*supra*) to produce phage particles presenting bla-bio fusion insertions. Those phage particles are again panned against streptavidin beads (*supra*). Additional shuffling rounds are carried out using selection conditions in which the ampicillin concentration is increased, and temperatures for growth, selection and panning are increased to 37°C. This allows the further optimization of the bla-bio insertion fusions with respect to activity, biotinylation level, folding and stability. The fusion(s) with optimal activity can be used for quantitation of streptavidin, e.g., by measuring beta-lactamase activity in a sandwich ELISA.

Table I

Primers, Oligonucleotides, Polynucleotides

GFP.For	AAGGAGATATACATATGGCTAGCAAAGGAGAAG
GFP.Rev	TTCACAGGTCAAGCTTCATTATTTGTAGAGCTCATC
Tet.For	TTAAGACCCACTTTCACATTTAAG
Tet.Rev	CTAAGCACTTGTCTCCTGTTTAC
Op1.For	CACTCTATCATTGATAGAGT
Op1.Rev	ACTCTATCAATGATAGAGTG
Op2.For	TCCCTATCAGTGATAGAGAA

Op1.Rev	TTCTCTATCACTGATAGGGA
Bla.For	TATTACTCGCGGCCCAGCCGGCCTTTGCTCACCCAGAAAC
Bla.Rev	TAGAATTCGGCCCCCGAGGCCAATGCTTAATCAGTGA
Bio.For	GGTTCTGAAGGTGGTGGTTCTGCTCAGCGTCTGTTCCACATCCTGG
	ACGCTCAGAAAATCGAATGGCACGGTCCGAAAGGTGGTTCTGGT
Bio.Rev	ACCAGAACCACCTTTCGGACCGTGCCATTCGATTTTCTGAGCGTCC
	AGGATGTGGAACAGACGCTGAGCAGAACCACCACCTTCAGAACC

***

Many modifications and variations of this invention can be made without departing from its spirit and scope, as will be apparent to those skilled in the art. The specific embodiments described herein are offered by way of example only, and the invention is to be limited only by the terms of the appended claims, along with the full scope of equivalents to which such claims are entitled.

All references cited herein are incorporated herein by reference in their entirety and for all purposes to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference in its entirety for all purposes.

WHAT IS CLAIMED IS:

1. A method of producing a DNA segment having a desired property or combination of properties, said method comprising:

a) mutating a substrate population, said substrate population comprising a plurality of DNA segments, wherein said mutating comprises

i) making insertions at random sites in said segments, or

ii) making deletions at random sites in said segments;

whereby a mutated population is produced, said mutated population comprising mutated DNA segments;

b) screening the mutated population to obtain a first selected population, said selected population comprising at least one DNA segment with a first desired property;

c) mutating the first selected population, wherein said mutating comprises

i) making insertions at random sites in the DNA segments in the selected population, or

ii) making deletions at random sites in the DNA segments in the selected population;

whereby a recursively mutated population is produced; and,

d) screening the recursively mutated population to obtain a recursively selected population, said recursively selected population comprising at least one DNA segment with a second desired property.

2. The method of claim 1, wherein the first desired property and the second desired property are the same.

3. The method of claim 2, wherein polynucleotides in the recursively selected population have a property that is enhanced when compared to the polynucleotides in the first selected population.

4. The method of claim 1, wherein the desired property is a combination of properties.

5. The method of claim 1, further comprising at least one additional cycle of mutation and screening after step (d), said cycle comprising mutating the recursively selected population and screening the resulting recursively mutated population to obtain new recursively selected population with a desired property.

6. The method of claim 1, wherein mutating in step (a) or step (c) comprises both making insertions and making deletions.

7. The method of claim 1, wherein the substrate population comprises DNA segments encoding a polypeptide or catalytic RNA.

8. The method of claim 7, wherein at least one screening step is for polynucleotides that encode a polypeptide having an activity selected from the group consisting of:

a) an enzymatic activity;

b) a substrate specificity; and,

c) a binding activity.

9. The method of claim 1, wherein the DNA segments comprise a promoter sequence.

10. The method of claim 1, wherein the DNA segments are vectors.

11. The method of claim 1, wherein the substrate population is homogeneous.

12. The method of claim 1, further comprising the step of shuffling one or a combination of polynucleotides in the recursively selected population.

13. The method of claim 5, further comprising the step of shuffling one or a combination of polynucleotides in the recursively selected population.

14. A method of producing a DNA segment having a desired property, said method comprising:

a) mutating a first substrate population, said substrate population comprising a plurality of DNA segments, wherein said mutating comprises

i) making insertions at random sites in said segments, or

ii) making deletions at random sites in said segments;

whereby a first mutated population of mutated DNA segments is produced;

b) mutating a second substrate population, said substrate population comprising a plurality of DNA segments, wherein said mutating comprises

i) making insertions at random sites in said segments, or

ii) making deletions at random sites in said segments;

whereby a second mutated population of mutated DNA segments is produced;

c) recombining the first substrate population and the second substrate population, whereby a recombined population is produced; and,

d) screening the recombined population to identify at least one DNA segment with the desired property.

15. The method of claim 14 wherein the first and second mutated populations are screened to produce a first and second selected population, each having a desired property, and the selected populations are recombined.

16. The method of claim 14, wherein the recombination is carried out by shuffling.

17. The method of claim 14, wherein the recombination is directed.

18. The method of claim 14, wherein the first desired property and the second desired property are the same.

19. The method of claim 14, wherein at least one screening step is for polynucleotides that encode a polypeptide having an activity selected from the group consisting of:

a) an enzymatic activity;

b) a substrate specificity; and,

c) a binding activity.

20. The method of claim 14, wherein the DNA segments comprise a promoter sequence.

21. The method of claim 14, wherein the DNA segments are vectors.

22. A method of producing a DNA segment having a desired property, said method comprising:

a) mutating a substrate population, said substrate population comprising a plurality of DNA segments, wherein said mutating comprises

i) making insertions at random sites in said segments, or

ii) making deletions at random sites in said segments;

whereby a mutated population is produced, said mutated population comprising mutated DNA segments;

b) screening the mutated population to obtain a selected population, said selected population comprising at least one DNA segment with the desired property;

c) shuffling at least one DNA segment for the selected population, whereby a recombined population is produced; and,

d) screening the recombined population for a desired property.

23. The method of claim 22, wherein the shuffling comprises conducting a polynucleotide amplification process on overlapping segments of at least one polynucleotide from the selected population under conditions whereby one segment serves as a template for extension of another segment, to generate a population of recombinant polynucleotides.

24. The method of claim 23, wherein at least one screening step is for polynucleotides that encode a polypeptide having an activity selected from the group consisting of:

a) an enzymatic activity;

b) a substrate specificity; and,

c) a binding activity.

25. The method of claim 23, wherein the DNA segments comprise a promoter sequence.

FIGURE 1

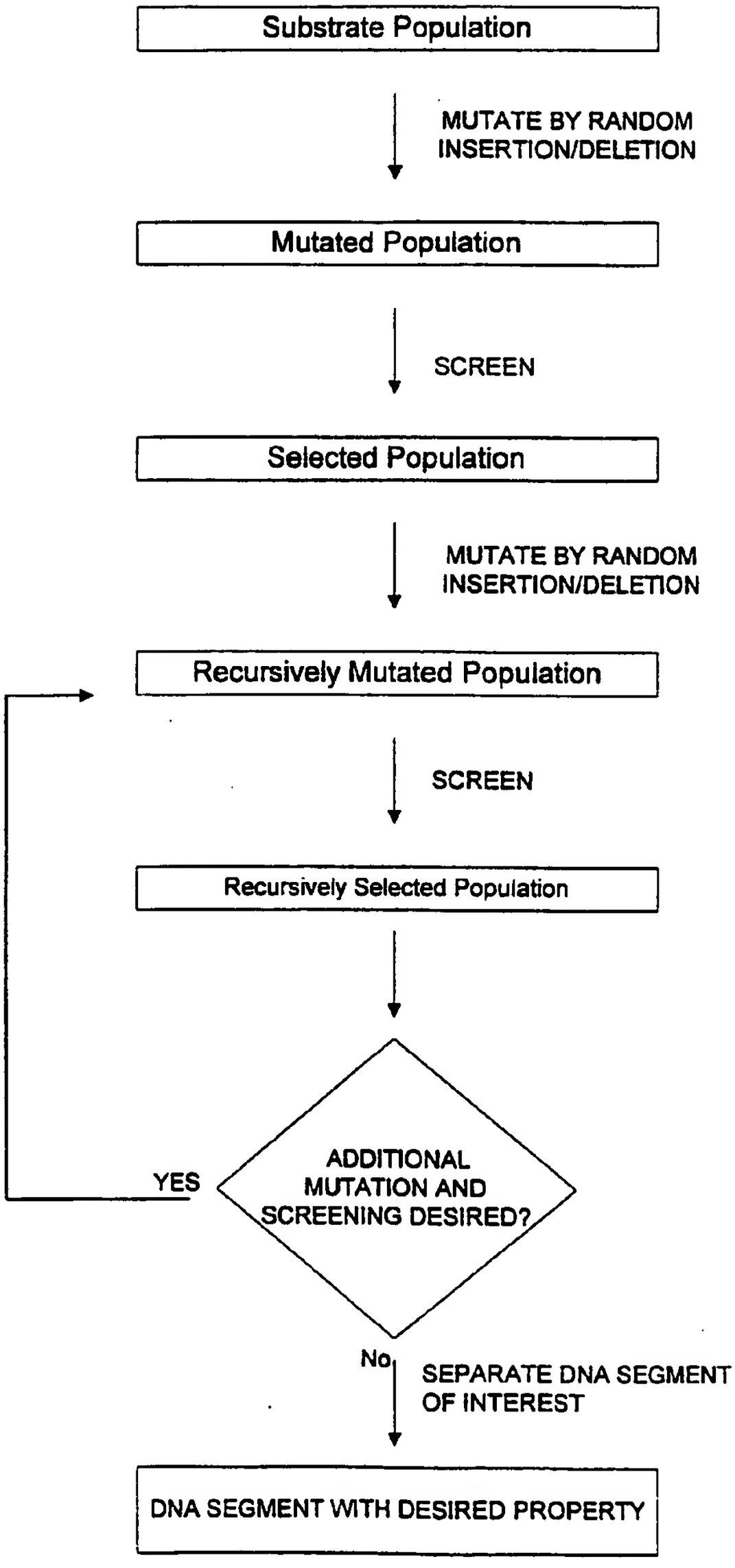

Substrate Population
MUTATE BY RANDOM INSERTION/DELETION
Mutated Population
SCREEN
Selected Population
MUTATE BY RANDOM INSERTION/DELETION
Recursively Mutated Population
SCREEN
Recursively Selected Population
ADDITIONAL MUTATION AND SCREENING DESIRED?
YES
No
SEPARATE DNA SEGMENT OF INTEREST
DNA SEGMENT WITH DESIRED PROPERTY

FIGURE 2

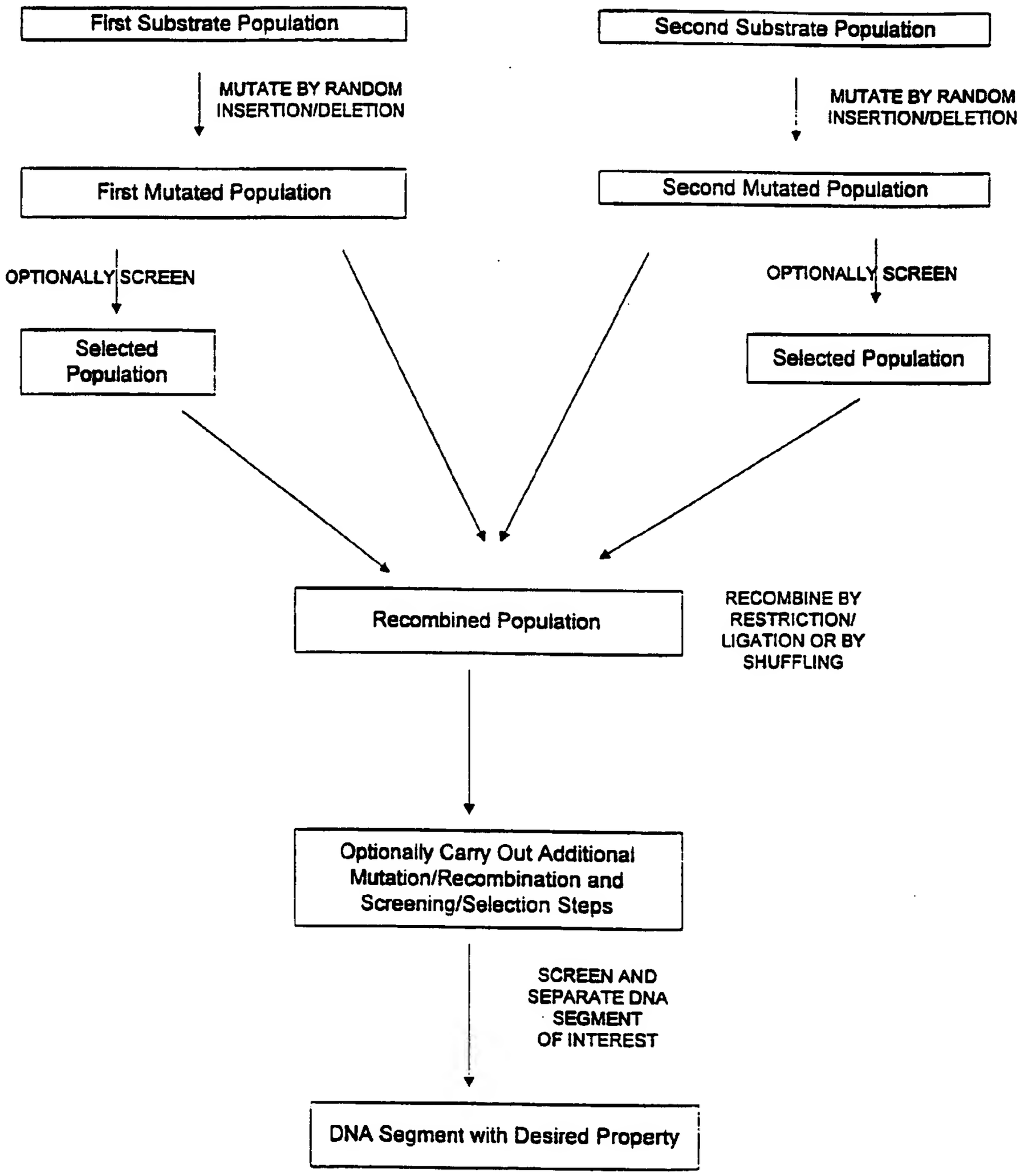
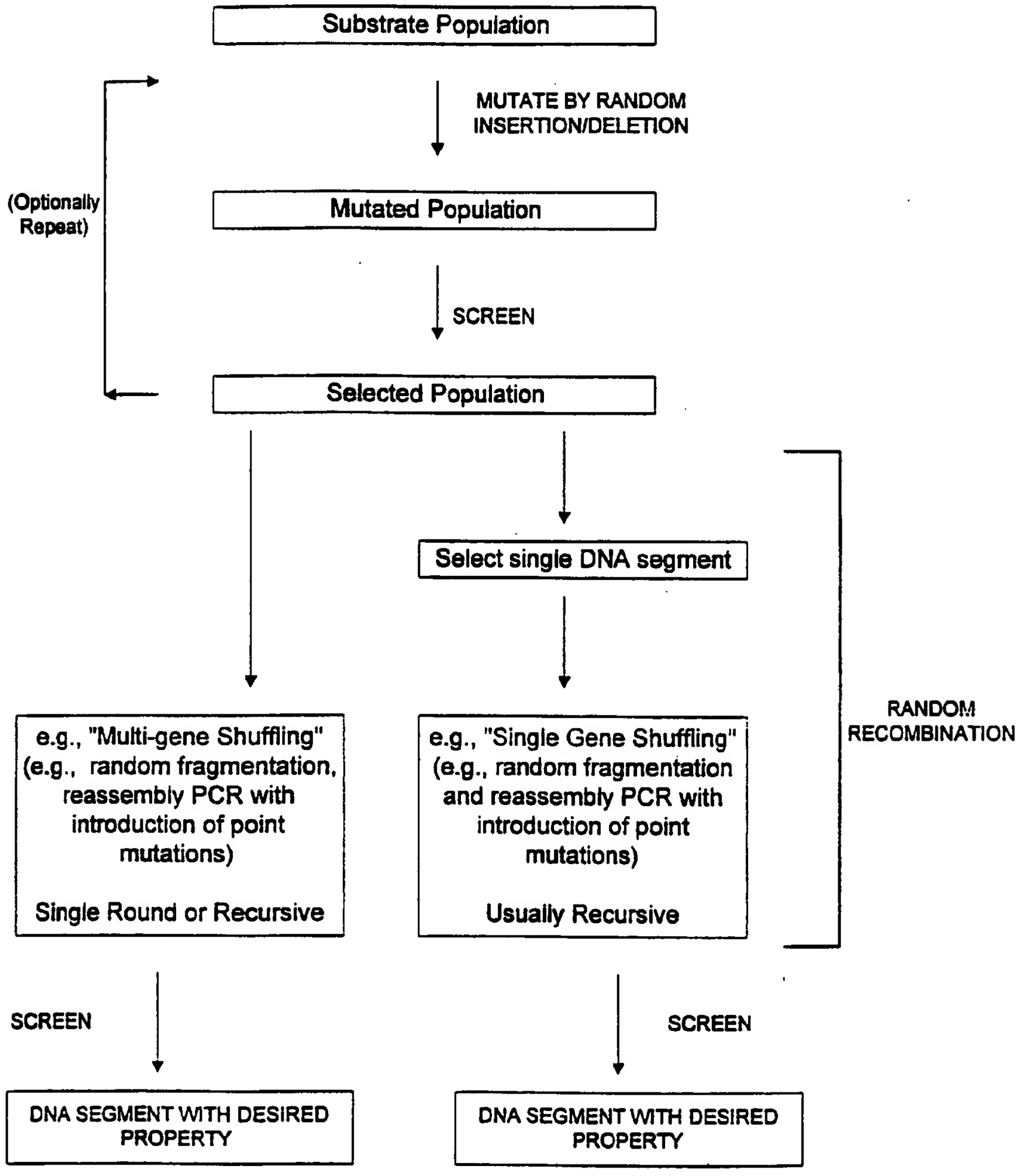

First Substrate Population
Second Substrate Population
MUTATE BY RANDOM INSERTION/DELETION
MUTATE BY RANDOM INSERTION/DELETION
First Mutated Population
Second Mutated Population
OPTIONALLY SCREEN
OPTIONALLY SCREEN
Selected Population
Selected Population
Recombined Population
RECOMBINE BY RESTRICTION/ LIGATION OR BY SHUFFLING
Optionally Carry Out Additional Mutation/Recombination and Screening/Selection Steps
SCREEN AND SEPARATE DNA SEGMENT OF INTEREST
DNA Segment with Desired Property

FIGURE 3


Substrate Population
MUTATE BY RANDOM INSERTION/DELETION
Mutated Population
(Optionally Repeat)
SCREEN
Selected Population
Select single DNA segment
RANDOM RECOMBINATION
e.g., "Multi-gene Shuffling" (e.g., random fragmentation, reassembly PCR with introduction of point mutations)
Single Round or Recursive
e.g., "Single Gene Shuffling" (e.g., random fragmentation and reassembly PCR with introduction of point mutations)
Usually Recursive
SCREEN
SCREEN
DNA SEGMENT WITH DESIRED PROPERTY
DNA SEGMENT WITH DESIRED PROPERTY

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12Q 1/68, C12P 19/34

A1

(11) International Publication Number: WO 99/29902

(43) International Publication Date: 17 June 1999 (17.06.99)

(21) International Application Number: PCT/US98/25698

(22) International Filing Date: 4 December 1998 (04.12.98)

(30) Priority Data:
60/067,908 8 December 1997 (08.12.97) US

(71) Applicant *(for all designated States except US)*: CALIFORNIA INSTITUTE OF TECHNOLOGY [US/US]; 1200 East California Boulevard, Pasadena, CA 91125 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: ARNOLD, Frances [US/US]; 629 South Grand Avenue, Pasadena, CA 91105 (US). SHAO, Zhixin [CN/DE]; Wohnung 52, Birkenstrasse 23, D–82377 Penzberg (DE). VOLKOV, Alexander [RU/–]; – (**).

(74) Agents: LIEBESCHUETZ, Joe et al.; Townsend and Townsend and Crew LLP, 8th floor, Two Embarcadero Center, San Francisco, CA 94111–3834 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHOD FOR CREATING POLYNUCLEOTIDE AND POLYPEPTIDE SEQUENCES

(57) Abstract

The invention provides methods for evolving a polynucleotide toward acquisition of a desired property. Such methods entail incubating a population of parental polynucleotide variants under conditions to generate annealed polynucleotides comprising heteroduplexes. The heteroduplexes are then exposed to a cellular DNA repair system to convert the heteroduplexes to parental polynucleotide variants or recombined polynucleotide variants. The resulting polynucleotides are then screened or selected for the desired property.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68, C12P 19/34

A1

(11) International Publication Number: **WO 99/29902**

(43) International Publication Date: 17 June 1999 (17.06.99)

(21) International Application Number: PCT/US98/25698

(22) International Filing Date: 4 December 1998 (04.12.98)

(30) Priority Data:
60/067,908 8 December 1997 (08.12.97) US

(71) Applicant *(for all designated States except US)*: CALIFORNIA INSTITUTE OF TECHNOLOGY [US/US]; 1200 East California Boulevard, Pasadena, CA 91125 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: ARNOLD, Frances [US/US]; 629 South Grand Avenue, Pasadena, CA 91105 (US). SHAO, Zhixin [CN/DE]; Wohnung 52, Birkenstrasse 23, D–82377 Penzberg (DE). VOLKOV, Alexander [RU/–]; – (**).

(74) Agents: LIEBESCHUETZ, Joe et al.; Townsend and Townsend and Crew LLP, 8th floor, Two Embarcadero Center, San Francisco, CA 94111–3834 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHOD FOR CREATING POLYNUCLEOTIDE AND POLYPEPTIDE SEQUENCES

(57) Abstract

The invention provides methods for evolving a polynucleotide toward acquisition of a desired property. Such methods entail incubating a population of parental polynucleotide variants under conditions to generate annealed polynucleotides comprising heteroduplexes. The heteroduplexes are then exposed to a cellular DNA repair system to convert the heteroduplexes to parental polynucleotide variants or recombined polynucleotide variants. The resulting polynucleotides are then screened or selected for the desired property.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

METHOD FOR CREATING POLYNUCLEOTIDE AND POLYPEPTIDE SEQUENCES

CROSS-REFERENCES TO RELATED APPLICATIONS

This application derives priority from USSN 60/067908, filed December 8, 1997, which is incorporated by reference in its entirety for all purposes.

TECHNICAL FIELD

The invention resides in the technical field of genetics, and more specifically, forced molecular evolution of polynucleotides to acquire desired properties.

BACKGROUND

A variety of approaches, including rational design and directed evolution, have been used to optimize protein functions (1, 2). The choice of approach for a given optimization problem depends, in part, on the degree of understanding of the relationships between sequence, structure and function. Rational redesign typically requires extensive knowledge of a structure-function relationship. Directed evolution requires little or no specific knowledge about structure-function relationship; rather, the essential features is a means to evaluate the function to be optimized. Directed evolution involves the generation of libraries of mutant molecules followed by selection or screening for the desired function. Gene products which show improvement with respect to the desired property or set of properties are identified by selection or screening. The gene(s) encoding those products can be subjected to further cycles of the process in order to accumulate beneficial mutations. This evolution can involve few or many generations, depending on how far one wishes to progress and the effects of mutations typically observed in each generation. Such approaches have been used to create novel functional nucleic acids (3, 4), peptides and other small molecules (3), antibodies (3), as well as enzymes and other proteins (5, 6, 7). These procedures are fairly tolerant to inaccuracies and noise in the function evaluation (7).

Several publications have discussed the role of gene recombination in directed evolution (see WO 97/07205, WO 98/42727, US 5807723, US 5,721,367, US 5,776,744 and WO 98/41645 US 5,811,238, WO 98/41622, WO 98/41623, and US 5,093,257).

A PCR-based group of recombination methods consists of DNA shuffling [5, 6], staggered extension process [89, 90] and random-priming recombination [87]. Such methods typically involve synthesis of significant amounts of DNA during assembly/recombination step and subsequent amplification of the final products and the efficiency of amplification decreases with gene size increase.

Yeast cells, which possess an active system for homologous recombination, have been used for *in vivo* recombination. Cells transformed with a vector and partially overlapping inserts efficiently join the inserts together in the regions of homology and restore a functional, covalently-closed plasmid [91]. This method does not require PCR amplification at any stage of recombination and therefore is free from the size considerations inherent in this method. However, the number of crossovers introduced in one recombination event is limited by the efficiency of transformation of one cell with multiple inserts. Other *in vivo* recombination methods entail recombination between two parental genes cloned on the same plasmid in a tandem orientation. One method relies on homologous recombination machinery of bacterial cells to produce chimeric genes [92]. A first gene in the tandem provides the N-terminal part of the target protein, and a second provides the C-terminal part. However, only one crossover can be generated by this approach. Another *in vivo* recombination method uses the same tandem organization of substrates in a vector [93]. Before transformation into *E. coli* cells, plasmids are linearized by endonuclease digestion between the parental sequences. Recombination is performed *in vivo* by the enzymes responsible for double-strand break repair. The ends of linear molecules are degraded by a 5'->3' exonuclease activity, followed by annealing of complementary single-strand 3' ends and restoration of the double-strand plasmid [94]. This method has similar advantages and disadvantages of tandem recombination on circular plasmid.

SUMMARY OF THE INVENTION

The invention provides methods for evolving a polynucleotide toward acquisition of a desired property. Such methods entail incubating a population of

parental polynucleotide variants under conditions to generate annealed polynucleotides comprises heteroduplexes. The heteroduplexes are then exposed to a cellular DNA repair system to convert the heteroduplexes to parental polynucleotide variants or recombined polynucleotide variants. The resulting polynucleotides are then screened or selected for the desired property.

In some methods, the heteroduplexes are exposed to a DNA repair system *in vitro*. A suitable repair system can be prepared in the form of cellular extracts.

In other methods, the products of annealing including heteroduplexes are introduced into host cells. The heteroduplexes are thus exposed to the host cells' DNA repair system in vivo.

In several methods, the introduction of annealed products into host cells selects for heteroduplexes relative to transformed cells comprising homoduplexes. Such can be achieved, for example, by providing a first polynucleotide variant as a component of a first vector, and a second polynucleotide variant is provided as a component of a second vector. The first and second vectors are converted to linearized forms in which the first and second polynucleotide variants occur at opposite ends. In the incubating step, single-stranded forms of the first linearized vector reanneal with each other to form linear first vector, single-stranded forms of the second linearized vector reanneal with each other to form linear second vector, and single-stranded linearized forms of the first and second vectors anneal with each to form a circular heteroduplex bearing a nick in each strand. Introduction of the products into cells thus selects for cirular heteroduplexes relative to the linear first and second vector. Optionally, in the above methods, the first and second vectors can be converted to linearized forms by PCR. Alternatively, the first and second vectors can be converted to linearized forms by digestion with first and second restriction enzymes.

In some methods, polynucleotide variants are provided in double stranded form and are converted to single stranded form before the annealing step. Optionally, such conversion is by conducting asymmetric amplification of the first and second double stranded polynucleotide variants to amplify a first strand of the first polynucleotide variant, and a second strand of the second polynucleotide variant. The first and second strands anneal in the incubating step to form a heteroduplex.

In some methods, a population of polynucleotides comprising first and second polynucleotides is provided in double stranded form, and the method further

comprises incorporating the first and second polynucleotides as components of first and second vectors, whereby the first and second polynucleotides occupy opposite ends of the first and second vectors. In the incubating step single-stranded forms of the first linearized vector reanneal with each other to form linear first vector, single-stranded forms of the second linearized vector reanneal with each other to form linear second vector, and single-stranded linearized forms of the first and second vectors anneal with each to form a circular heteroduplex bearing a nick in each strand. In the introducing step selects for transformed cells comprises the circular heteroduplexes relative to the linear first and second vector.

In some methods, the first and second polynucleotides are obtained from chromosomal DNA. In some methods, the polynucleotide variants encode variants of a polypeptide. In some methods, the population of polynucleotide variants comprises at least 20 variants. In some methods, the population of polynucleotide variants are at least 10 kb in length.

In some methods, the polynucleotide variants comprises natural variants. In other methods, the polynucleotide variants comprise variants generated by mutagenic PCR or cassette mutagenesis. In some methods, the host cells into which heteroduplexes are introduced are bacterial cells. In some methods, the population of variant polynucleotide variants comprises at least 5 polynucleotides having at least 90% sequence identity with one another.

Some methods further comprise a step of at least partially demethylating variant polynucleotides. Demethylation can be achieved by PCR amplification or by passaging variants through methylation-deficient host cells.

Some methods include a further step of sealing one or more nicks in heteroduplex molecules before exposing the heteroduplexes to a DNA repair system. Nicks can be sealed by treatment with DNA ligase.

Some methods further comprise a step of isolating a screened recombinant polynucleotide ariant. In some methods, the polynucleotide variant is screened to produce a recombinant protein or a secondary metabolite whose production is catalyzed thereby.

In some methods, the recombinant protein or secondary metabolite is formulated with a carrier to form a pharmaceutical composition.

In some methods, the polynucleotide variants encode enzymes selected from the group consisting of proteases, lipases, amylases, cutinases, cellulases, amylases, oxidases, peroxidases and phytases. In other methods, the polynucleotide variants encode a polypeptide selected from the group consisting of insulin, ACTH, glucagon, somatostatin, somatotropin, thymosin, parathyroid hormone, pigmentary hormones, somatomedin, erthropoietin, luteinizing hormone, chorionic gonadotropin, hyperthalmic releasing factors, antidiuretic hormones, thyroid stimulating hormone, relaxin, interferon, thrombopoietic (TPO), and prolactin.

In some methods, each polynucleotide in the population of variant polynucleotides encodes a plurality of enzymes forming a metabolic pathway.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates the process of heteroduplex formation using polymerase chain reaction (PCR) with one set of primers for each different sequence to amplify the target sequence and vector.

Figure 2 illustrates the process of heteroduplex formation using restriction enzymes to linearize the target sequences and vector.

Figure 3 illustrates a process of heteroduplex formation using asymmetric or single primer polymerase chain reaction (PCR) with one set of primers for each different sequence to amplify the target sequence and vector.

Figure 4 illustrates heteroduplex recombination using unique restriction enzymes (X and Y) to remove the homoduplexes.

Figure 5 shows the amino acid sequences of the FlaA from *R. lupini* (SEQ ID NO: 1) and *R. meliloti* (SEQ ID NO:2).

Figures 6A and 6B show the locations of the unique restriction sites utilized to linearize pRL20 and pRM40.

Figures 7A, B, C and D show the DNA sequences of four mosaic *flaA* genes created by *in vitro* heteroduplex formation followed by *in vivo* repair ((a) is SEQ ID NO:3, (b) is SEQ ID NO:4, (c) is SEQ ID NO:5 and (d) is SEQ ID NO:6).

Figure 8 illustrates how the heteroduplex repair process created mosaic *flaA* genes containing sequence information from both parent genes.

Figure 9 shows the physical maps of *Actinoplanes utahensis* ECB deacylase mutants with enhanced specific activity ((a) is pM7-2 for Mutant 7-2, and (b) is pM16 for Mutant 16).

Figure 10 illustrates the process used for Example 2 to recombine mutations in Mutant 7-2 and Mutant 16 to yield ECB deacylase recombinant with more enhanced specific activity.

Figure 11 shows specific activities of wild-type ECB deacylase and improved mutants Mutant 7-2, Mutant 16 and recombined Mutant 15.

Figure 12 shows positions of DNA base changes and amino acid substitutions in recombined ECB deacylase Mutant 15 with respect to parental sequences of Mutant 7-2 and Mutant 16.

Figures 13 A, B, C, D and E show the DNA sequence of *A.utahensis* ECB deacylase gene mutant M-15 genes created by *in vitro* heteroduplex formation followed by *in vivo* repair (SEQ ID NO:7).

Figure 14 illustrates the process used for Example 3 to recombine mutations in RC1 and RC2 to yield thermostable subtilisin E.

Figure 15 illustrates the sequences of RC1 and RC2 and the ten clones picked randomly from the transformants of the reaction products of duplex formation as described in Example 3. The x's correspond to base positions that differ between RC1 and RC2. The mutation at 995 corresponds to amino acid substitution at 181, while that at 1107 corresponds to an amino acid substitution at 218 in the subtilisin protein sequence.

Figure 16 shows the results of screening 400 clones from the library created by heteroduplex formation and repair for initial activity (A_i) and residual activity (A_r). The ratio A_i/ A_r was used to estimate the enzymes' thermostability. Data from active variants are sorted and plotted in descending order. Approximately 12.9% of the clones exhibit a phenotype corresponding to the double mutant containing both the N181D and the N218S mutations.

DEFINITIONS

Screening is, in general, a two-step process in which one first physically separates the cells and then determines which cells do and do not possess a desired property. Selection is a form of screening in which identification and physical separation are achieved simultaneously by expression of a selection marker, which, in some genetic

circumstances. allows cells expressing the marker to survive while other cells die (or vice versa). Exemplary screening members include luciferase, βgalactosidase and green fluorescent protein. Selection markers include drug and toxin resistance genes. Although spontaneous selection can and does occur in the course of natural evolution. in the present methods selection is performed by man.

An exogenous DNA segment is one foreign (or heterologous) to the cell or homologous to the cell but in a position within the host cell nucleic acid in which the element is not ordinarily found. Exogenous DNA segments are expressed to yield exogenous polypeptides.

The term gene is used broadly to refer to any segment of DNA associated with a biological function. Thus, genes include coding sequences and/or the regulatory sequences required for their expression. Genes also include nonexpressed DNA segments that, for example, form recognition sequences for other proteins.

The term "wild-type" means that the nucleic acid fragment does not comprise any mutations. A "wild-type" protein means that the protein will be active at a level of activity found in nature and typically will comprise the amino acid sequence found in nature. In an aspect, the term "wild type" or "parental sequence" can indicate a starting or reference sequence prior to a manipulation of the invention.

"Substantially pure" means an object species is the predominant species present (i.e., on a molar basis it is more abundant than any other individual macromolecular species in the composition), and preferably a substantially purified fraction is a composition wherein the object species comprises at least about 50 percent (on a molar basis) of all macromolecular species present. Generally, a substantially pure composition will comprise more than about 80 to 90 percent of all macromolecular species present in the composition. Most preferably, the object species is purified to essential homogeneity (contaminant species cannot be detected in the composition by conventional detection methods) wherein the composition consists essentially of a single macromolecular species. Solvent species, small molecules (<500 Daltons). and elemental ion species are not considered macromolecular species.

Percentage sequence identity is calculated by comparing two optimally aligned sequences over the window of comparison. determining the number of positions at which the identical nucleic acid base occurs in both sequences to yield the number of matched positions. dividing the number of matched positions by the total number of

positions in the window of comparison. Optimal alignment of sequences for aligning a comparison window can be conducted by computerized implementations of algorithms GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, WI.

The term naturally-occurring is used to describe an object that can be found in nature as distinct from being artificially produced by man. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory is naturally-occurring. Generally, the term naturally-occurring refers to an object as present in a non-pathological (undiseased) individual, such as would be typical for the species.

A nucleic acid is operably linked when it is placed into a functional relationship with another nucleic acid sequence. For instance, a promoter or enhancer is operably linked to a coding sequence if it increases the transcription of the coding sequence. Operably linked means that the DNA sequences being linked are typically contiguous and, where necessary to join two protein coding regions, contiguous and in reading frame. However, since enhancers generally function when separated from the promoter by several kilobases and intronic sequences may be of variable lengths, some polynucleotide elements may be operably linked but not contiguous.

A specific binding affinity between, for example, a ligand and a receptor, means a binding affinity of at least $1 \times 10^6\ M^{-1}$.

The term "cognate" as used herein refers to a gene sequence that is evolutionarily and functionally related between species. For example but not limitation, in the human genome, the human CD4 gene is the cognate gene to the mouse CD4 gene, since the sequences and structures of these two genes indicate that they are highly homologous and both genes encode a protein which functions in signaling T cell activation through MHC class II-restricted antigen recognition.

The term "heteroduplex" refers to hybrid DNA generated by base pairing between complementary single strands derived from the different parental duplex molecules, whereas the term "homoduplex" refers to double-stranded DNA generated by base pairing between complementary single strands derived from the same parental duplex molecules.

The term "nick" in duplex DNA refers to the absence of a phosphodiester bond between two adjacent nucleotides on one strand. The term "gap" in duplex DNA refers to an absence of one or more nucleotides in one strand of the duplex. The term "loop" in duplex DNA refers to one or more unpaired nucleotides in one strand.

A mutant or variant sequence is a sequence showing substantial variation from a wild type or reference sequence that differs from the wild type or reference sequence at one or more positions.

DETAILED DESCRIPTION

I. General

The invention provides methods of evolving a polynucleotide toward acquisition of a desired property. The substrates for the method are a population of at least two polynucleotide variant sequences that contain regions of similarity with each other but, which also have point(s) or regions of divergence. The substrates are annealed *in vitro* at the regions of similarity. Annealing can regenerate initial substrates or can form heteroduplexes, in which component strands originate from different parents. The products of annealing are exposed to enzymes of a DNA repair, and optionally a replication system, that repairs unmatched pairings. Exposure can be *in vivo* as when annealed products are transformed into host cells and exposed to the hosts DNA repair system. Alternatively, exposure can be *in vitro*, as when annealed products are exposed to cellular extracts containing functional DNA repair systems. Exposure of heteroduplexes to a DNA repair system results in DNA repair at bulges in the heteroduplexes due to DNA mismatching. The repair process differs from homologous recombination in promoting nonreciprocal exchange of diversity between strands. The DNA repair process is typically effected on both component strands of a heteroduplex molecule and at any particular mismatch is typically random as to which strand is repaired. The resulting population can thus contain recombinant polynucleotides encompassing an essentially random reassortment of points of divergence between parental strands. The population of recombinant polynucleotides is then screened for acquisition of a desired property. The property can be a property of the polynucleotide *per se*, such as capacity of a DNA molecule to bind to a protein or can be a property of an expression product thereof. such as mRNA or a protein.

II. Substrates For Shuffling

The substrates for shuffling are variants of a reference polynucleotide that show some region(s) of similarity with the reference and other region(s) or point(s) of divergence. Regions of similarity should be sufficient to support annealing of polynucleotides such that stable heteroduplexes can be formed. Variants forms often show substantial sequence identity with each other (e.g., at least 50%, 75%, 90% or 99%). There should be at least sufficient diversity between substrates that recombination can generate more diverse products than there are starting materials. Thus, there must be at least two substrates differing in at least two positions. The degree of diversity depends on the length of the substrate being recombined and the extent of the functional change to be evolved. Diversity at between 0.1-25% of positions is typical. Recombination of mutations from very closely related genes or even whole sections of sequences from more distantly related genes or sets of genes can enhance the rate of evolution and the acquisition of desirable new properties. Recombination to create chimeric or mosaic genes can be useful in order to combine desirable features of two or more parents into a single gene or set of genes, or to create novel functional features not found in the parents. The number of different substrates to be combined can vary widely in size from two to 10, 100, 1000, to more than 10^5, 10^7, or 10^9 members.

The initial small population of the specific nucleic acid sequences having mutations may be created by a number of different methods. Mutations may be created by error-prone PCR. Error-prone PCR uses low-fidelity polymerization conditions to introduce a low level of point mutations randomly over a long sequence. Alternatively, mutations can be introduced into the template polynucleotide by oligonucleotide-directed mutagenesis. In oligonucleotide-directed mutagenesis, a short sequence of the polynucleotide is removed from the polynucleotide using restriction enzyme digestion and is replaced with a synthetic polynucleotide in which various bases have been altered from the original sequence. The polynucleotide sequence can also be altered by chemical mutagenesis. Chemical mutagens include, for example, sodium bisulfite, nitrous acid, hydroxylamine, hydrazine or formic acid. Other agents which are analogues of nucleotide precursors include nitrosoguanidine, 5-bromouracil, 2-aminopurine, or acridine. Generally, these agents are added to the PCR reaction in place of the nucleotide precursor thereby mutating the sequence. Intercalating agents such as proflavine.

acriflavine, quinacrine and the like can also be used. Random mutagenesis of the polynucleotide sequence can also be achieved by irradiation with X-rays or ultraviolet light. Generally, plasmid DNA or DNA fragments so mutagenized are introduced into E. coli and propagated as a pool or library of mutant plasmids.

Alternatively the small mixed population of specific nucleic acids can be found in nature in the form of different alleles of the same gene or the same gene from different related species (i.e., cognate genes). Alternatively, substrates can be related but nonallelic genes, such as the immunoglobulin genes. Diversity can also be the result of previous recombination or shuffling. Diversity can also result from resynthesizing genes encoding natural proteins with alternative codon usage.

The starting substrates encode variant forms of sequences to be evolved. In some methods, the substrates encode variant forms of a protein for which evolution of a new or modified property is desired. In other methods, the substrates can encode variant forms of a plurality of genes constituting a multigene pathway. In such methods, variation can occur in one or any number of the component genes. In other methods, substrates can contain variants segments to be evolved as DNA or RNA binding sequences. In methods, in which starting substrates containing coding sequences, any essential regulatory sequences, such as a promoter and polyadenylation sequence, required for expression may also be present as a component of the substrate. Alternatively, such regulatory sequences can be provided as components of vectors used for cloning the substrates.

The starting substrates can vary in length from about 50, 250, 1000, 10,000, 100,000, 10^6 or more bases. The starting substrates can be provided in double- or single-stranded form. The starting substrates can be DNA or RNA and analogs thereof. If DNA, the starting substrates can be genomic or cDNA. If the substrates are RNA, the substrates are typically reverse-transcribed to cDNA before heteroduplex formation. Substrates can be provided as cloned fragments, chemically synthesized fragments or PCR amplification products. Substrates can derive from chromosomal, plasmid or viral sources. In some methods, substrates are provided in concatemeric form.

III. Procedures for Generating Heteroduplexes

Heteroduplexes are generated from double stranded DNA substrates, by denaturing the DNA substrates and incubating under annealing conditions. Hybridization

conditions for heteroduplex formation are sequence-dependent and are different in different circumstances. Longer sequences hybridize specifically at higher temperatures. Generally, hybridization conditions are selected to be about 25°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH, and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium.

Exemplary conditions for denaturation and renaturation of double stranded substrates are as follows. Equimolar concentrations (~ 1.0 - 5.0 nM) of the substrates are mixed in 1 x SSPE buffer (180 mM NaCl, 1.0 mM EDTA, 10 mM NaH_2PO_4, pH 7.4) After heating at 96°C for 10 minutes, the reaction mixture is immediately cooled at 0°C for 5 minutes; The mixture is then incubated at 68°C for 2-6 hr. Denaturation and reannealing can also be carried out by the addition and removal of a denaturant such as NaOH. The process is the same for single stranded DNA substrates, except that the denaturing step may be omitted for short sequences.

By appropriate design of substrates for heteroduplex formation, it is possible to achieve selection for heteroduplexes relative to reformed parental homoduplexes. Homoduplexes merely reconstruct parental substrates and effectively dilute recombinant products in subsequent screening steps. In general, selection is achieved by designing substrates such that heteroduplexes are formed in open-circles, whereas homoduplexes are formed as linear molecules. A subsequent transformation step results in substantial enrichment (e.g., 100-fold) for the circular heteroduplexes.

Figure 1 shows a method in which two substrate sequences in separate vectors are PCR-amplified using two different sets of primers (P1, P2 and P3, P4). Typically, first and second substrates are inserted into separate copies of the same vector. The two different pairs of primers initiate amplification at different points on the two vectors. Fig. 1 shows an arrangement in which the P1/P2 primer pairs initiates amplification at one of the two boundaries of the vector with the substrate and the P1/P2 primer pair initiates replication at the other boundary in a second vector. The two primers in each primer pair prime amplification in opposite directions around a circular plasmid. The amplification products generated by this amplification are double-stranded linearized vector molecules in which the first and second substrates occur at opposite ends of the

vector. The amplification products are mixed, denatured and annealed. Mixing and denaturation can be performed in either order. Reannealing generates two linear homoduplexes, and an open circular heteroduplex containing one nick in each strand, at the initiation point of PCR amplification. Introduction of the amplification products into host cells selects for the heteroduplexes relative to the homoduplexes because the former transform much more efficiently than the latter.

It is not essential in the above scheme that amplification is initiated at the interface between substrate and the rest of the vector. Rather, amplification can be initiated at any points on two vectors bearing substrates provided that the amplification is initiated at different points between the vectors. In the general case, such amplification generates two linearized vectors in which the first and second substrates respectively occupy different positions relative to the remainder of the vector. Denaturation and reannealing generator heteroduplexes similar to that shown in Fig. 1, except that the nicks occur within the vector component rather than at the interface between plasmid and substrate. Initiation of amplification outside the substrate component of a vector has the advantage that it is not necessary to design primers specific for the substrate borne by the vector.

Although Fig. 1 is exemplified for two substrates, the above scheme can be extended to any number of substrates. For example, an initial population of vector bearing substrates can be divided into two pools. One pool is PCR-amplified from one set of primers, and the other pool from another. The amplification products are denatured and annealed as before. Heteroduplexes can form containing one strand from any substrate in the first pool and one strand from any substrate in the second pool. Alternatively, three or more substrates cloned into multiple copies of a vector can be subjected to amplification with amplification in each vector starting at a different point. For each substrate, this process generates amplification products varying in how flanking vector DNA is divided on the two sides of the substrate. For example, one amplification product might have most of the vector on one side of the substrate, another amplification product might have most of the vector on the other side of the substrate, and a further amplification product might have an equal division of vector sequence flanking the substrate. In the subsequent annealing step, a strand of substrate can form a circular heteroduplex with a strand of any other substrate, but strands of the same substrate can only reanneal with each other to form a linear homoduplex. In a still further variation,

multiple substrates can be performed by performing multiple iterations of the scheme in Fig. 1. After the first iteration, recombinant polynucleotides in a vector, undergo heteroduplex formation with a third substrate incorporated into a further copy of the vector. The vector bearing the recombinant polynucleotides and the vector bearing the third substrate are separately PCR amplified from different primer pairs. The amplification products are then denatured and annealed. The process can be repeated further times to allow recombination with further substrates.

An alternative scheme for heteroduplex formation is shown in Fig. 2. Here, first and second substrates are incorporated into separate copies of a vector. The two copies are then respectively digested with different restriction enzymes. Fig. 2 shows an arrangement in which, the restriction enzymes cut at opposite boundaries between substrates and vector, but all that is necessary is to use two different restriction enzymes that cut at different places. Digestion generates linearized first and second vector bearing first and second substrates, the first and second substrates occupying different positions relative to the remaining vector sequences. Denaturation and reannealing generates open circular heteroduplexes and linear homoduplexes. The scheme can be extended to recombination between more than two substrates using analogous strategies to those described with respect to Fig. 1. In one variation, two pools of substrates are formed, and each is separately cloned into vector. The two pools are then cute with different enzymes, and annealing proceeds as for two substrates. In another variation, three or more substrates can be cloned into three or more copies of vector, and the three or more result molecules cut with three or more enzymes, cutting at three or more sites. This generates three different linearized vector forms differing in the division of vector sequences flanking the substrate moiety in the vectors. Alternatively, any number of substrates can be recombined pairwise in an iterative fashion with products of one round of recombination annealing with a fresh substrate in each round.

In a further variation, heteroduplexes can be formed from substrates molecules in vector-free form, and the heteroduplexes subsequently cloned into vectors. Such can be achieved by asymmetric amplification of first and second substrates as shown in Fig. 3. Asymmetric or single primer PCR amplifies only one strand of a duplex. By appropriate selection of primers, opposite strands can be amplified from two different substrates. On reannealing amplification products, heteroduplexes are formed from opposite strands of the two substrates. Because only one strand is amplified from each

substrate, reannealing does not reform homoduplexes (other than for small quantities of unamplified substrate). The process can be extended to allow recombination of any number of substrates using analogous strategies to those described with respect to Fig. 1. For example, substrates can be divided into two pools. and each pool subject to the same asymmetric amplification, such that amplification products of one pool can only anneal with amplification products of the other pool , and not with each other. Alternatively, shuffling can proceed pairwise in an iterative manner, in which recombinants formed from heteroduplexes of first and second substrates, are subsequently subjected to heteroduplex formation with a third substrate. Point mutations can also be introduced at a desired level during PCR amplification.

Fig. 4 shows another approach of selecting for heteroduplexes relative to homoduplexes. First and second substrates are isolated by PCR amplification from separate vectors. The substrates are denatured and allowed to anneal forming both heteroduplexes and reconstructed homoduplexes. The products of annealing are digested with restriction enzymes X and Y. X has a site in the first substrate but not the second substrate, and vice versa for Y. Enzyme X cuts reconstructed homoduplex from the first substrate and enzyme Y cuts reconstructed homoduplex from the second substrate. Neither enzyme cuts heteroduplexes. Heteroduplexes can effectively be separated from restriction fragments of homoduplexes by further cleavage with enzymes A and B having sites proximate to the ends of both the first and second substrates, and ligation of the products into vector having cohesive ends compatible with ends resulting from digestion with A and B. Only heteroduplexes cut with A and B can ligate with the vector. Alternatively, heteroduplexes can be separated from restriction fragments of homoduplexes by size selection on gels. The above process can be generalized to N substrates by cleaving the mixture of heteroduplexes and homoduplexes with N enzymes, each one of which cuts a different substrate and no other substrate. Heteroduplexes can be formed by directional cloning. Two substrates for heteroduplex formation can be obtained by PCR amplification of chromosomal DNA and joined to opposite ends of a linear vector. Directional cloning can be achieved by digesting the vector with two different enzymes, and digesting or adapting first and second substrates to be respectively compatible with cohesive ends of only of the two enzymes used to cut the vector. The first and second substrates can thus be ligated at opposite ends of a linearized vector fragment. This scheme can be extended to any number of substrates by using principles

analogous to those described for Fig. 1. For example, substrates can be divided into two pools before ligation to the vector. Alternatively, recombinant products formed by heteroduplex formation of first and second substrates, can subsequently undergo heteroduplex formation with a third substrate.

IV. Vectors and Transformation

In general, substrates are incorporated into vectors either before or after the heteroduplex formation step. A variety of cloning vectors typically used in genetic engineering are suitable.

The vectors containing the DNA segments of interest can be transferred into the host cell by standard methods, depending on the type of cellular host. For example, calcium chloride transformation is commonly utilized for prokaryotic cells, whereas calcium phosphate treatment. Lipofection, or electroporation may be used for other cellular hosts. Other methods used to transform mammalian cells include the use of Polybrene, protoplast fusion, liposomes, electroporation, and microinjection, and biolisitics (see, generally, Sambrook et al., supra). Viral vectors can also be packaged *in vitro* and introduced by infection. The choice of vector depends on the host cells. In general, a suitable vector has an origin of replication recognized in the desired host cell, a selection maker capable of being expressed in the intended host cells and/or regulatory sequences to support expression of genes within substrates being shuffled.

V. Types of Host Cells

In general any type of cells supporting DNA repair and replication of heteroduplexes introduced into the cells can be used. Cells of particular interest are the standard cell types commonly used in genetic engineering, such as bacteria, particularly, *E. coli (16, 17)*. Suitable *E. coli* strains include *E. coli mutS, mutL, dam*$^-$, and/or *recA*$^+$, *E.coli* XL-10-Gold ([*Tet*^{r} *Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte] [F' proAB lacI*^{q}*ZΔM15 Tn10 (Tet*^{r}*) Amy Cam*^{r}*]*), *E. coli* ES1301 *mutS* [Genotype: *lacZ53, mutS201::Tn5, thyA36, rha-5, metB1, deoC, IN(rrnD-rrnE)*] (20, 24, 28-42). Preferred E. coli strains are *E.coli* SCS110 [Genotype: *rpsl (Str*^{r}*), thr, leu, enda, thi-1, lacy, galk, galt, ara tona, tsx, dam, dcm, supE44, Δ(lac-proAB), [F, traD36, proA*$^+$*B*$^+$*lacI*^{q}*ZΔM15*], which have normal cellular mismatch repair

systems (17). This strain type repairs mismatches and unmatches in the heteroduplex with little strand-specific preference. Further, because this strain is *dam*⁻ and *dcm*⁻, plasmid isolated from the strain is unmethylated and therefore particularly amenable for further rounds of DNA duplex formation/mismatch repair (see below). Other suitable bacterial cells include gram-negative and gram-positive, such as *Bacillus. Pseudomonas,* and *Salmonella.*

Eukaryotic organisms are also able to carry out mismatch repair (43-48). Mismatch repair systems in both prokaryotes and eukaryotes are thought to play an important role in the maintenance of genetic fidelity during DNA replication. Some of the genes that play important roles in mismatch repair in prokaryotes, particularly *mutS* and *mutL*, have homologs in eukaryotes. in the outcome of genetic recombinations, and in genome stability. Wild-type or mutant *S. cerevisiae* has been shown to carry out mismatch repair of heteroduplexes (49-56), as have COS-1 monkey cells (57). Preferred strains of yeast are Picchia and Saccharomyces. Mammalian cells have been shown to have the capacity to repair G-T to G-C base pairs by a short-patch mechanism (38, 58-63). Mammalian cells (e.g., mouse, hamster, primate, human), both cell lines and primary cultures can also be used. Such cells include stem cells, including embryonic stem cells, zygotes, fibroblasts, lymphocytes, Chinese hamster ovary (CHO), mouse fibroblasts (NIH3T3), kidney, liver, muscle, and skin cells. Other eucaryotic cells of interest include plant cells, such as maize, rice, wheat, cotton, soybean, sugarcane, tobacco. and arabidopsis; fish, algae. fungi (aspergillus, podospora. neurospora), insect (e.g., baculo lepidoptera) (see, Winnacker, "From Genes to Clones," VCH Publishers. N.Y., (1987), which is incorporated herein by reference).

In vivo repair occurs in a wide variety of prokaryotic and eukaryotic cells. Use of mammalian cells is advantage in certain application in which substrates encode polypeptides that are expressed only in mammalian cells or which are intended for use in mammalian cells. However, bacterial and yeast cells are advantageous for screening large libraries due to the higher transformation frequencies attainable in these strains.

V. In Vitro DNA Repair Systems

As an alternative to introducing annealed products into host cells, annealed products can be exposed a DNA repair system *in vitro*. The DNA repair system can be obtained as extracts from repair-competent *E. coli*, yeast or any other cells (64-67).

Repair-competent cells are lysed in appropriate buffer and supplemented with nucleotides. DNA is incubated in this cell extract and transformed into competent cells for replication.

VI. Screening and Selection

After introduction of annealed products into host cells, the host cells are typically cultured to allow repair and replication to occur and optionally, for genes encoded by polynucleotides to be expressed. The recombinant polynucleotides can be subject to further rounds of recombination using the heteroduplex procedures described above, or other shuffling methods described below. However, whether after one cycle of recombination or several. recombinant polynucleotides are subjected to screening or selection for a desired property. In some instances, screening or selection in performed in the same host cells that are used for DNA repair. In other instances, recombinant polynucleotides. their expression products or secondary metabolites produced by the expression products are isolated from such cells and screened in vitro. In other instances, recombinant polynucleotides are isolated from the host cells in which recombination occurs and are screened or selected in other host cells. For example, in some methods, it is advantageous to allow DNA repair to occur in a bacterial host strain, but to screen an expression product of recombinant polynucleotides in eucaryotic cells. The recombinant polynucleotides surviving screening or selection are sometimes useful products in themselves. In other instances, such recombinant polynucleotides are subjected to further recombination with each other or other substrates. Such recombination can be effected by the heteroduplex methods described above or any other shuffling methods. Further round(s) of recombination are followed by further rounds of screening or selection on an iterative basis. Optionally, the stringency of selection can be increased at each round.

The nature of screening or selection depends on the desired property sought to be acquired. Desirable properties of enzymes include high catalytic activity, capacity to confer resistance to drugs, high stability, the ability to accept a wider (or narrower) range of substrates. or the ability to function in nonnatural environments such as organic solvents. Other desirable properties of proteins include capacity to bind a selected target. secretion capacity, capacity to generate an immune response to a given target, lack of immunogenicity and toxicity to pathogenic microorganisms. Desirable properties of DNA or RNA polynucleotides sequences include capacity to specifically

bind a given protein target, and capacity to regulate expression of operably linked coding sequences. Some of the above properties. such as drug resistance. can be selected by plating cells on the drug. Other properties. such as the influence of a regulatory sequence on expression, can be screened by detecting appearance of the expression product of a reporter gene linked to the regulatory sequence. Other properties, such as capacity of an expressed protein to be secreted, can be screened by FACS™, using a labelled antibody to the protein. Other properties, such as immunogenicity or lack thereof. can be screened by isolating protein from individual cells or pools of cells, and analyzing the protein *in vitro* or in a laboratory animal.

VII. Variations

1. Demethylation

Most cell types methylate DNA in some manner. with the pattern of methylation differing between cells types. Sites of methylation include 5-methylcytosine (m^5C), N4-methylcytosine (m^4C) and N^6-methyladenine (m^6A), 5-hydroxymethylcytosine (hm^5C) and 5-hydroxymethyluracil (hm^5U). In *E. coli*, methylation is effected by *Dam* and *Dcm* enzymes. The methylase specified by the *dam* gene methylates the N6-position of the adenine residue in the sequence GATC, and the methylase specified by the *dcm* gene methylates the C5-position of the internal cytosine residue in the sequence CCWGG. DNA from plants and mammal is often subject to CG methylation meaning that CG or CNG sequences are methylated. Possible effects of methylated on cellular repair are discussed by references 18-20.

In some methods, DNA substrates for heteroduplex formation are at least partially demethylated on one or both strands, preferably the latter. Demethylation of substrate DNA promotes efficient and random repair of the heteroduplexes. In heteroduplexes formed with one strand dam-methylated and one strand unmethylated, repair is biased to the unmethylated strand. with the methylated strand serving as the template for correction. If neither strand is methylated. mismatch repair occurrs. but showes insignificant strand preference (23, 24).

Demethylation can be performed in a variety of ways. In some methods, substrate DNA is demethylated by PCR-amplification. In some instances. DNA demethylation is accomplished in one of the PCR steps in the heteroduplex formation

procedures described above. In other methods. an additional PCR step is performed to effect demethylation. In other methods, demethylation is effected by passaging substrate DNA through methylation deficient host cells (e.g. an *E. coli dam*⁻*dcm*⁻ strain). In other methods, substrate DNA is demethylated *in vitro* using a demethylating enzymes. Demethylated DNA is used for heteroduplex formation using the same procedures described above. Heteroduplexes are subsequently introduced into DNA-repair-proficient but restriction-enzyme-defective cells to prevent degradation of the unmethylated heteroduplexes.

2. Sealing Nicks

Several of the methods for heteroduplex formation described above result in circular heteroduplexes bearing nicks in each strand. These nicks can be sealed before introducing heteroduplexes into host cells. Sealing can be effected by treatment with DNA ligase under standard ligating conditions. Ligation forms a phosphodiester bond to link two adjacent bases separated by a nick in one strand of double helix of DNA. Sealing of nicks increases the frequency of recombination after introduction of heteroduplexes into host cells.

3. Error Prone PCR Attendant To Amplification

Several of the formats described above include a PCR amplification step. Optionally, such a step can be performed under mutagenic conditions to induce additional diversity between substrates.

VIII. Other Shuffling Methods

The methods of heteroduplex formation described above can be used in conjunction with other shuffling methods. For example. one can perform one cycle of heteroduplex shuffling, screening or selection, followed by a cycle of shuffling by another method, followed by a further cycle of screening or selection. Other shuffling formats are described by WO 95/22625: US 5,605.793; US 5.811.238; WO 96/19256; Stemmer. *Science* 270. 1510 (1995); Stemmer et al.. *Gene*. 164. 49-53 (1995); Stemmer. *Bio/Technology*, 13. 549-553 (1995); Stemmer. *Proc. Natl. Acad. Sci.* USA 91. 10747-10751 (1994); Stemmer. *Nature* 370. 389-391 (1994); Crameri et al., *Nature Medicine*. 2(1):1-3. (1996); Crameri et al.. *Nature Biotechnology* 14. 315-319 (1996);WO 98/42727;

WO 98/41622; WO 98/05764 and WO 98/42728. WO 98/27230 (each of which is incorporated by reference in its entirety for all purposes).

IX. Protein Analogs

Proteins isolated by the methods also serve as lead compounds for the development of derivative compounds. The derivative compounds can include chemical modifications of amino acids or replace amino acids with chemical structures. The analogs should have a stabilized electronic configuration and molecular conformation that allows key functional groups to be presented in substantially the same way as a lead protein. In particular, the non-peptic compounds have spatial electronic properties which are comparable to the polypeptide binding region, but will typically be much smaller molecules than the polypeptides, frequently having a molecular weight below about 2 CHD and preferably below about 1 CHD. Identification of such non-peptic compounds can be performed through several standard methods such as self-consistent field (CSF) analysis, configuration interaction (CHI) analysis, and normal mode dynamics analysis. Computer programs for implementing these techniques are readily available. See Rein et al., *Computer-Assisted Modeling of Receptor-Ligand Interactions* (Alan Liss, New York, 1989).

IX. Pharmaceutical Compositions

Polynucleotides, their expression products, and secondary metabolites whose formation is catalyzed by expression products, generated by the above methods are optionally formulated as pharmaceutical compositions. Such a composition comprises one or more active agents, and a pharmaceutically acceptable carrier. A variety of aqueous carriers can be used, e.g., water, buffered water, phosphate-buffered saline (PBS), 0.4% saline, 0.3% glycine, human albumin solution and the like. These solutions are sterile and generally free of particulate matter. The compositions may contain pharmaceutically acceptable auxiliary substances as required to approximate physiological conditions such as pH adjusting and buffering agents, toxicity adjusting agents and the like, for example, sodium acetate, sodium chloride, potassium chloride, calcium chloride and sodium is selected primarily based on fluid volumes, viscosities, and so forth, in accordance with the particular mode of administration selected.

EXAMPLES

EXAMPLE 1. Novel *Rhizobium Flaa* Genes From Recombination Of *Rhizobium Lupini Flaa* And *Rhizobium Meliloti FlaA*

Bacterial flagella have a helical filament. a proximal hook and a basal body with the flagellar motor (68). This basic design has been extensively examined in *E. coli* and *S. typhimurium* and is broadly applicable to many other bacteria as well as some archaea. The long helical filaments are polymers assembled from flagellin subunits, whose molecular weights range between 20,000 and 65,000, depending on the bacterial species (69). Two types of flagellar filaments, named plain and complex, have been distinguished by their electron microscopically determined surface structures (70). Plain filaments have a smooth surface with faint helical lines, whereas complex filaments exhibit a conspicuous helical pattern of alternating ridges and grooves. These characteristics of complex flagellar filaments are considered to be responsible for the brittle and (by implication) rigid structure that enables them to propel bacteria efficiently in viscous media (71-73). Whereas flagella with plain filaments can alternate between clockwise and counter clockwise rotation (68), all known flagella with complex filaments rotate only clockwise with intermittent stops (74). Since this latter navigation pattern is found throughout bacteria and archaea, it has been suggested that complex flagella may reflect the common background of an ancient, basic motility design (69).

Differing from plain bacterial flagella in the fine structure of their filaments dominated by conspicuous helical bands and in their fragility, the filaments are also resistant against heat decomposition (72). Schmitt et al. (75) showed that bacteriophage 7-7-1 specifically adsorbs to the complex flagella of *R.lupini* H 13-3 and requires motility for a productive infection of its host. Though the flagellins from *R. meliloti* and *R. lupini* are quite similar, bacteriophage 7-7-1 does not infect *R.meliloti*. Until now complex flagella have been observed in only three species of soil bacteria: *Pseudomonas rhodos* (73), *R.meliloti* (76), and *R.lupini* H13-3 (70, 72). Cells of *R.lupini* H13-3 posses 5 to 10 peritrichously inserted complex flagella. which were first isolated and analyzed by high resolution electron microscopy and by optical diffraction (70).

Maruyama et al. (77) further found that a higner content of hydrophobic amino acid residues in the complex filament may be one of the main reasons for the unusual properties of complex flagella. By measuring mass per unit length and obtaining three-dimensional reconstruction from electron micrographs. Trachtenberg et al. (73. 78)

suggested that the complex filaments of *R. lupini* are composed of functional dimers. Figure 6 shows the comparison between the deduced amino acid sequence of the *R. lupini* H13-3 FlaA and the deduced amino acid sequence of the *R. meliloti* FlaA. Perfect matches are indicated by vertical lines, and conservative exchanges are indicated by colons. The overall identity is 56%. The *R.lupini flaA* and *R.meliloti flaA* were subjected to *in vitro* heteroduplex formation followed by *in vivo* repair in order to create novel FlaA molecules and structures.

A. Methods

pRL20 containing *R. lupini* H13-3 *flaA* gene and pRM40 containing *R.meliloti flaA* gene are shown in Figs. 6A and 6B. These plasmids were isolated from *E. coli* SCS110 (free from dam- and dcm-type methylation).

About 3.0 pg. of unmethylated pRL20 and pRM40 DNA were digested with *Bam* HI and *Eco* RI, respectively, at 37°C for 1 hour. After agarose gel separation, the linearized DNA was purified with Wizard PCR Prep kit (Promega, WI, USA).

Equimolar concentrations (2.5 nM) of the linearized unmethylated pRL20 and pRM40 were mixed in 1 x SSPE buffer (180 mM NaCl, 1 mM EDTA, 10 mM NaH2PO4, pH 7.4). After heating at 96°C for 10 minutes, the reaction mixture was immediately cooled at 0°C for 5 minutes. The mixture was incubated at 68°C for 2 hour for heteroduplexes to form.

One microliter of the reaction mixture was used to transform 50 µl of *E. coli* ES 1301 *mutS*, *E. coli* SCS110 and *E. coli* JM109 competent cells. The transformation efficiency with *E. coli* JM109 competent cells was about seven times higher than that of *E. coli* SCS110 and ten times higher than that of *E. coli* ES1301 *mutS*, although the overall transformation efficiencies were 10-200 times lower than those of control transformations with the close, covalent and circular pUC19 plasmid.

Two clones were selected at random from the *E. coli* SCS110 transformants and two from *E. coli* ES1301 mutS transformants, and plasmid DNA was isolated from these four clones for further DNA sequencing analysis.

B. Results

Figure 7 shows (a) the sequence of SCS01 (clone#1 from *E. coli* SCS110 transformant library), (b) the sequence of SCS02 (clone #2 from *E. coli* SCS110

transformant library), (c) the sequence of ES01 (clone #1 from *E. coli* ES1301 transformant library), and (d) the sequence of ES02 (clone #2 from *E. coli* ES1301 transformant library). All four sequences were different from wild-type *R. lupini flaA* and *R. meliloti flaA* sequences. Clones SCS02. ES01 and ES02 all contain a complete open-reading frame, but SCS01 was truncated. Figure 8 shows that recombination mainly occurred in the loop regions (unmatched regions). The *flaA* mutant library generated from *R. meliloti flaA* and *R. lupini flaA* can be transformed into *E. coli* SCS110, ES1301, XL10-Gold and JM109, and transformants screened for functional FlaA recombinants.

EXAMPLE 2. Directed Evolution Of ECB Deacylase For Variants With Enhanced Specific Activity

Streptomyces are among the most important industrial microorganisms due to their ability to produce numerous important secondary metabolites (including many antibiotics) as well as large amounts of enzymes. The approach described here can be used with little modification for directed evolution of native *Streptomyces* enzymes, some or all of the genes in a metabolic pathways, as well as other heterologous enzymes expressed in *Streptomyces*.

New antifungal agents are critically needed by the large and growing numbers of immune-compromised AIDS, organ transplant and cancer chemotherapy patients who suffer opportunistic infections. Echinocandin B (ECB), a lipopeptide produced by some species of *Aspergillus*. has been studied extensively as a potential antifungal. Various antifungal agents with significantly reduced toxicity have been generated by replacing the linoleic acid side chain of *A. nidulans* echinocandin B with different aryl side chains (79-83). The cyclic hexapeptide ECB nucleus precursor for the chemical acylation is obtained by enzymatic hydrolysis of ECB using *Actinoplanes utahensis* ECB deacylase. To maximize the conversion of ECB into intact nucleus. this reaction is carried out at pH 5.5 with a small amount of miscible organic solvent to solubilize the ECB substrate. The product cyclic hexapeptide nucleus is unstable at pH above 5.5 during the long incubation required to fully deacylate ECB (84). The pH optimum of ECB deacylase. however. is 8.0-8.5 and its activity is reduced at pH 5.5 and in the presence of more than 2.5% ethanol (84). To improve production of ECB nucleus it is necessary to increase the activity of the ECB deacylase under these process-relevant conditions.

Relatively little is known about ECB deacylase. The enzyme is a heterodimer whose two subunits are derived by processing of a single precursor protein (83). The 19.9 kD α-subunit is separated from the 60.4 kD β-subunit by a 15-amino acid spacer peptide that is removed along with a signal peptide and another spacer peptide in the native organism. The polypeptide is also expressed and processed into functional enzyme in *Streptomyces lividans*, the organism used for large-scale conversion of ECB by recombinant ECB deacylase. The three-dimensional structure of the enzyme has not been determined, and its sequence shows so little similarity to other possibly related enzymes such as penicillin acylase that a structural model reliable enough to guide a rational effort to engineer the ECB deacylase will be difficult to build. We therefore decided to use directed evolution (85) to improve this important activity.

Protocols suitable for mutagenic PCR and random-priming recombination of the 2.4 kb ECB deacylase gene (73% G+C) have been described recently (86). Here, we further describe the use of heteroduplex recombination to generate new ECB deacylase with enhanced specific activity.

In this case, two *Actinoplanes utahensis* ECB deacylase mutants, M7-2 and M16, which show higher specific activity at pH 5.5 and in the presence of 10% MeOH were recombined using technique of the *in vitro* heteroduplex formation and *in vivo* mismatch repair .

Figure 12 shows the physical maps of plasmids pM7-2 and pM16 which contain the genes for the M7-2 and M16 ECB deacylase mutants. Mutant M7-2 was obtained through mutagenic PCR performed directly on whole *Streptomyces lividans* cells containing wild-type ECB deacylase gene, expressed from plasmid pSHP150-2*. *Streptomyces* with pM7-2 show 1.5 times the specific activity of cells expressing the wild-type ECB deacylase (86). Clone pM16 was obtained using the random-priming recombination technique as described (86, 87). It shows 2.4 times specific activity of the wild-type ECB deacylase clone.

A. Methods:

M7-2 and M16 plasmid DNA (pM7-2 and pM16) (Fig. 9) were purified from *E. coli* SCS110 (in separate reactions). About 5.0 µg of unmethylated M7-2 and M16 DNA were digested with *Xho* I and *Psh* AI. respectively, at 37°C for 1 hour (Fig.

10). After agarose gel separation. the linearized DNA was purified using a Wizard PCR Prep Kit (Promega, WI. USA).

Equimolar concentrations (2.0 nM) of the linearized unmethylated pM7-2 and pM16 DNA were mixed in 1 x SSPE buffer (1x SSPE: 180 mM NaCl, 1.0 mM EDTA, 10 mM NaH_2PO_4, pH 7.4). After heating at 96 °C for 10 minutes, the reaction mixture is immediately cooled at 0 °C for 5 minutes. The mixture was incubated at 68 °C for 3 hours to promote formation of heteroduplexes.

One microliter of the reaction mixture was used to transform 50 µl of *E.coli* ES1301 *mutS*, SCS110 and JM109 competent cells. All transformants from *E. coli* ES1301 *mutS* were pooled and *E. coli* SCS110 were pooled. A plasmid pool was isolated from each pooled library, and this pool was used to transform *S. lividans* TK23 protoplasts to form a mutant library for deacylase activity screening.

Transformants from the *S. lividans* TK23 libraries were screened for ECB deacylase activity with an *in situ* plate assay. Transformed protoplasts were allowed to regenerate on R2YE agar plates for 24 hr at 30°C and to develop in the presence of thiostrepton for 48 hours. When the colonies grew to the proper size, 6 ml of 0.7% agarose solution containing 0.5 mg/ml ECB in 0.1 M sodium acetate buffer (pH 5.5) was poured on top of each R2YE-agar plate and allowed to develop for 18-24 hr at 30°C. Colonies surrounded by a clearing zone larger than that of a control colony containing wild-type plasmid pSHP150-2*, were selected for further characterization.

Selected transformants were inoculated into 20 ml medium containing thiostrepton and grown aerobically at 30°C for 48 hours, at which point they were analyzed for ECB deacylase activity using HPLC. 100 µl of whole broth was used for a reaction at 30 °C for 30 minutes in 0.1 M NaAc buffer (pH 5.5) containing 10% (v/v) MeOH and 200 µg/ml of ECB substrate. The reactions were stopped by adding 2.5 volumes of methanol. and 20 µl of each sample were analyzed by HPLC on a 100 x 4.6 mm polyhydroxyethyl aspartamide column (PolyLC Inc., Columbia, MD, USA) at room temperature using a linear acetonitrile gradient starting with 50:50 of A:B (A = 93% acetonitrile. 0.1% phosphoric acid: B = 70% acetonitrile. 0.1% phosphoric acid) and ending with 30:70 of A:B in 22 min at a flow rate of 2.2 ml/min. The areas of the ECB and ECB nucleus peaks were calculated and subtracted from the areas of the

corresponding peaks from a sample culture of *S. lividans* containing pIJ702* in order to estimate the ECB deacylase activity.

2.0 ml pre-cultures of positive mutants were used to inoculate 50-ml medium and allowed to grow at 30°C for 96 hr. The supernatants were further concentrated to 1/30 their original volume using an Amicon filtration unit (Beverly, MA, USA) with molecular weight cutoff of 10 kD. The resulting enzyme samples were diluted with an equal volume of 50 mM KH_2PO4 (pH 6.0) buffer and were applied to Hi-Trap ion exchange column (Pharmacia Biotech, Piscataway, NJ, USA) . The binding buffer was 50 mM KH_2PO_4 (pH 6.0), and the elution buffer was 50 mM KH_2PO_4 (pH 6.0) containing 1.0 M NaCl. A linear gradient from 0 to 1.0 M NaCl was applied in 8 column volumes with a flow rate of 2.7 ml/min. The ECB deacylase fraction eluting at 0.3 M NaCl was concentrated and the buffer was exchanged for 50 mM KH_2PO_4 (pH 6.0) using Centricon-10 units. Enzyme purity was verified by SDS-PAGE using Coomassie Blue stain, and the concentration was determined using the Bio-Rad Protein Assay Reagent (Hercules, CA, USA).

A modified HPLC assay was used to determine the activities of the ECB deacylase mutants on ECB substrate (84). Four μg of each purified ECB deacylase mutant was used for activity assay reaction at 30°C for 30 minutes in 0.1 M NaAc buffer (pH 5.5) containing 10% (v/v) MeOH and different concentrations of ECB substrate. Assays were performed in duplicate. The reactions were stopped by adding 2.5 volumes of methanol, and the HPLC assays were carried out as described above. The absorbance values were recorded, and the initial rates were calculated by least-squares regression of the time progress curves from which the Km and the kcat were calculated.

Activities as a function of pH were measured for the purified ECB deacylases at 30°C at different pH values: 5, 5.5 and 6 (0.1 M acetate buffer); 7, 7.5, 8 and 8.5 (0.1 M phosphate buffer); 9 and 10 (0.1 M carbonate buffer) using the HPLC assay. Stabilities of purified ECB deacylases were were determined at 30°C in 0.1 M NaAc buffer (pH 5.5) containing 10% methanol. Samples were withdrawn at different time intervals. and the residual activity was measured in the same buffer with the HPLC assay described above.

B. Results

Fig. 11 shows that after one round of applying this heteroduplex repair technique on the mutant M7-2 and M16 genes. one mutant (M15) from about 500 original transformants was found to possess 3.1 times the specific activity of wild-type.

Wild type and evolved M15 ECB deacylases were purified and their kinetic parameters for deacylation of ECB were determined by HPLC. The evolved deacylases M15 has an increased catalytic rate constant, k_{cat} by 205%. The catalytic efficiency (k_{cat}/K_m) of M20 is enhanced by a factor of 2.9 over the wild-type enzyme.

Initial rates of deacylation with the wild type and M15 at different pH values from 5 to 10 were determined at 200 μg/ml of ECB. The recombined M15 is more active than wild type at pH 5-8. Although the pH dependence of the enzyme activity in this assay is not strong, there is a definite shift of 1.0-1.5 units in the optimum to lower pH, as compared to wild type.

The time courses of deactivation of the purified ECB deacylase mutant M15 was measured in 0.1 M NaAc (pH 5.5) at 30°C. No significant difference in stability was observed between wild type and mutant M15.

The DNA mutations with respect to the wild type ECB deacylase sequence and the positions of the amino acid substitutions in the evolved variants M7-2, M16 and M15 are summarized in Figure 12.

The heteroduplex recombination technique can recombine parent sequences to create novel progeny. Recombination of the M7-2 and M16 genes yielded M15, whose activity is higher than any of its parents (Fid. 13). Of the six base substitutions in M15, five (at positions α50, α171, β57, β129 and β340) were inherited from M7-2, and the other one (β30) came from M16.

This approach provides an alternative to existing methods of DNA recombination and is particularly useful in recombining large genes or entire operons. This method can be used to create recombinant proteins to improve their properties or to study structure-function relationship.

EXAMPLE 3. Novel Thermostable *Bacillus Subtilis* Subtilisin E Variants

This example demonstrates the use *in vitro* heteroduplex formation followed by *in vivo* repair for combining sequence information from two different sequences in order to improve the thermostability of *Bacillus subtilis* subtilisin E.

Genes RC1 and RC2 encode thermostable *B. sublilis* subtilisin E variants (88). The mutations at base positions 1107 in RC1 and 995 in RC2 (Figure 14), giving rise to amino acid substitutions Asn218/Ser (N218S) and Asn181/Asp (N181 ID), lead to improvements in subtilisin E thermostability; the remaining mutations, both synonymous and nonsynonymous, have no detectable effects on thermostability. At 65°C, the single variants Nl8lD and N218S have approximately 3-fold and 2-fold longer half-lives, respectively, than wild subtilisin E. and variants containing both mutations have half-lives that are 8-fold longer (88). The different half-lives in a population of subtilisin E variants can therefore be used to estimate the efficiency by which sequence information is combined. In particular. recombination between these two mutations (in the absence of point mutations affecting thermostability) should generate a library in which 25% of the population exhibits the thermos/ability of the double mutant. Similarly, 25% of the population should exhibit wild-type like stability, as N181D and N218S are eliminated at equal frequency. We used the fractions of the recombined population as a diagnostic

A. Methods

The strategy underlying this example is shown in Fig. 15.

Subtilisin E thermostable mutant genes RC1 and RC2 (Fig. 14) are 986-bp fragments including 45 nt of subtilisin E prosequence, the entire mature sequence and 113 nt after the stop codon. The genes were cloned between *Bam* HI and *Nde* I in *E. coli/B. subtilis* shuttle vector pBE3. resulting in pBE3-1 and pBE3-2, respectively. Plasmid DNA pBE3-1 and pBE3-2 was isolated from *E.coli* SCS110.

About 5.0 µg of ummethylated pBE3-1 and pBE3-2 DNA were digested with *Bam* HI and *Nde* I. respectively. at 37°C for 1 hour. After agarose gel separation. equimolar concentrations (2.0nM) of the linearized unmethylated pBE3-1 and pBE3-2 were mixed in 1 x SSPE buffer (180 mM NaCl. 1.0 mM EDTA. 10 mM NaH_2PO_4. pH 7.4). After heating at 96°C for 10 minutes. the reaction mixture was immediately cooled at 0°C for 5 min. The mixture was incubated at 68°C for 2 hr for heteroduplexes to form.

One microliter of the reaction mixture was used to transform 50 µl of *E. coli* ES 1301 *mutS*, *E. coli* SCS110 and *E. coli* HB101 competent cells.

The transformation efficiency with *E. coli* HB101 competent cells was about ten times higher than that of *E.coli* SCS110 and 15 times higher than that of *E. coli* ES1301 *mutS*. But in all these cases, the transformation efficiencies were 10-250 times lower than that of the transformation with closed, covalent and circular control pUC19 plasmids.

Five clones from *E. coli* SCS110 mutant library and five from *E. coli* ES1301 *mutS* library were randomly chosen, and plasmid DNA was isolated using a QIAprep spin plasmid miniprep kit for further DNA sequencing analysis.

About 2.000 random clones from *E. coli* HB101 mutant library were pooled and total plasmid DNA was isolated using a QIAGEN-100 column. 0.5-4.0 ug of the isolated plasmid was used to transform *Bacillus subtilis* DB428 as described previously (88).

About 400 transformants from the *Bacillus subtilis* DB428 library were subjected to screening. Screening was performed using the assay described previously (88), on succinyl-Ala-Ala-Pro-Phe-p-nitroanilide. *B. subtilis* DB428 containing the plasmid library were grown on LB plates containing kanamycin (20 µg/ml) plates. After 18 hours at 37°C single colonies were picked into 96-well plates containing 200 µl SG/kanamycin medium per well. These plates were incubated with shaking at 37°C for 24 hours to let the cells to grow to saturation. The cells were spun down, and the supernatants were sampled for the thermostability assay.

Two replicates of 96-well assay plates were prepared for each growth plate by transferring 10 µl of supernatant into the replica plates. The subtilisin activities were then measured by adding 100 µl of activity assay solution (0.2 mM succinyl-Ala-Ala-Pro-Phe-p-nitroanilide, 100 mM Tris-HCl, 10 mM $CaCl_2$, pH 8.0, 37°C). Reaction velocities were measured at 405 nm to over 1.0 min in a ThermoMax microplate reader (Molecular Devices, Sunnyvale CA). Activity measured at room temperature was used to calculate the fraction of active clones (clones with activity less than 10% of that of wild type were scored as inactive). Initial activity (A_i) was measured after incubating one assay plate at 65°C for 10 minutes by immediately adding 100 µl of prewarmed (37°C) assay solution (0.2 mM succinyl-Ala-Ala-Pro-Phe-p-nitroanilide, 100 mM Tris-HCl, pH 8.0, 10 mM

$CaCl_2$, pH 8.0) into each well. Residual activity (Ar) was measured after 40 minute incubation.

B. Results

In vitro heteroduplex formation and *in vivo* repair was carried out as described above. Five clones from *E. coli* SCS110 mutant library and five from *E. coli* ES1301 *mutS* libraries were selected at random and sequenced. Fig. 14 shows that four out of the ten clones were different from the parent genes. The frequency of occurrence of a particular point mutation from parent RC1 or RC2 in the resulting genes ranged from 0% to 50%, and the ten point mutations in the heteroduplex have been repaired without strong strand-specific preference.

Since none of the ten mutations locates within the dcm site, the mismatch repair appears generally done via the *E. coli* long-patch mismatch repair systems. The system repairs different mismatches in a strand-specific manner using the state of N6-methylation of adenine in GATC sequences as the major mechanism for determining the strand to be repaired. With heteroduplexes methylated at GATC sequences on only one DNA strand, repair was shown to be highly biased to the unmethylated strand, with the methylated strand serving as the template for correction. If neither strand was methylated, mismatch repair occurred, but showed little strand preference (23, 24). These results shows that it is preferable to demethylate the DNA to be recombined to promote efficient and random repair of the heteroduplexes.

The rates of subtilisin E thermo-inactivation at 65°C were estimated by analyzing the 400 random clones from the *Bacillus subtilis* DB428 library. The thermostabilities obtained from one 96-well plate are shown in Figure 16, plotted in descending order. About 12.9% of the clones exhibited thermostability comparable to the mutant with the NI8lD and N218S double mutations. Since this rate is only half of that expected for random recombination of these two markers, it indicates that the two mismatches at positions 995 and 1107 within the heteroduplexes have been repaired with lower position randomness.

Sequence analysis of the clone exhibiting the highest thermostability among the screened 400 transformants from the *E. coli* SCS110 heteroduplex library confirmed the presence of both N181D and N218S mutations. Among the 400 transformants from the *B.sublilis* DB428 library that were screened, approximately 91%

of the clones expressed N181D- and/or N218S-type enzyme stabilities, while about 8.0% of the transformants showed only wild-type subtilisin E stability.

Less than 1.0% inactive clone was found, indicating that few new point mutations were introduced in the recombination process. This is consistent with the fact that no new point mutations were identified in the ten sequenced genes (Figure 14). While point mutations may provide useful diversity for some *in vitro* evolution applications, they can also be problematic for recombination of beneficial mutations, especially when the mutation rate is high.

EXAMPLE 4. Optimizing Conditions For The Heteroduplex Recombination.

We have found that the efficiency of heteroduplex recombination can differ considerably from gene to gene [17,57]. In this example, we investigate and optimize a variety of parameters that improve recombination efficiency.

DNA substrates used in this example were site-directed mutants of green fluorescent protein from *Aequorea victoria*. The GFP mutants had a stop codon(s) introduced at different locations along the sequence that abolished their fluorescence. Fluorescent wild type protein could be only restored by recombination between two or more mutations. Fraction of fluorescent colonies was used as a measure of recombination efficiency.

A. Methods

About 2-4 μg of each parent plasmid was used in one recombination experiment. One parent plasmid was digested with *Pst* I endonuclease another parent with *Eco*RI. Linearized plasmids were mixed together and 20 x SSPE buffer was added to the final concentration 1x (180 mM NaCl, 1 mM EDTA, 10 mM NaH_2PO_4, pH 7.4). The reaction mixture was heated at 96°C for 4 minutes, immediately transferred on ice for 4 minutes and the incubation was continued for 2 hours at 68°C.

Target genes were amplified in a PCR reaction with primers corresponding to the vector sequence of pGFP plasmid. Forward primer: 5'-CCGACTGGAAAGCGGGCAGTG-3', reverse primer 5'-CGGGGCTGGCTTAACTATGCGG-3'. PCR products were mixed together and purified using Qiagen PCR purification kit. Purified products were mixed with 20 x SSPE buffer and hybridized as described above. Annealed products were precipitated with ethanol or

purified on Qiagen columns and digested with *Eco*RI and *Pst*I enzymes. Digested products were ligated into *Pst*I and *Eco*RI digested pGFP vector.

dUTP was added into PCR reaction at final concentrations 200 μM, 40μM, 8 μM, 1.6 μM, 0.32 μM. PCR reaction and subsequent cloning procedures were performed as described above.

Recombinant plasmids were transformed into XL10 *E. coli* strain by a modified chemical transformation method. Cells were plated on ampicillin containing LB agar plates and grown overnight at 37°C, followed by incubation at room temperature or at 4°C until fluorescence developed.

B. Results.

1. Effect of ligation on recombination efficiency.

Two experiments have been performed to test the effect of breaks in the DNA heteroduplex on the efficiency of recombination. In one experiment heteroduplex plasmid was treated with DNA ligase to close all existing single-strand breaks and was transformed in identical conditions as an unligated sample (see Table 1). The ligated samples show up to 7-fold improvement in recombination efficiency over unligated samples.

In another experiment, dUTP was added into PCR reaction to introduce additional breaks into DNA upon repair by uracyl N-glycosylase in the host cells. Table 2 shows that dUMP incorporation significantly suppressed recombination, the extent of suppression increasing with increased dUTP concentration.

2. Effect of plasmid size on the efficiency of heteroduplex formation.

Plasmid size was a significant factor affecting recombination efficiency. Two plasmids pGFP (3.3 kb) and a *Bacillus* shuttle vector pCT1 (about 9 kb) were used in preparing circular heteroduplex-like plasmids following traditional heteroduplex protocol. For the purpose of this experiment (to study the effect of plasmid size on duplex formation), both parents had the same sequences. While pGFP formed about 30-40% of circular plasmid, the shuttle vector yielded less than 10% of this form.

Increase in plasmid size decreases concentration of the ends in the vicinity of each and makes annealing of very long (>0.8 kb) ends that are single-stranded more difficult. This difficulty is avoided by the procedure shown in Fig. 3, in which heteroduplex formation occurs between substrates in vector-free form, and heteroduplexes are subsequently inserted into a vector.

3. Efficiency of Recombination vs. Distance Between Mutations

A series of GFP variants was recombined pairwise to study the effect of distance between mutations on the efficiency of recombination. Parental genes were amplified by PCR, annealed and ligated back into pGFP vector. Heteroduplexes were transformed into XL10 *E.coli* strain.

The first three columns in Table 3 show the results of three independent experiments and demonstrate the dependence of recombination efficiency on the distance between mutations. As expected recombination becomes less and less efficient for very close mutations.

However, it is still remarkable that long-patch repair has been able to recombine mutations separated by only 27 bp.

The last line in Table 3 represents recombination between one single and one double mutants. Wild type GFP could only be restored in the event of double crossover with each individual crossover occurring in the distance of 99 bp only, demonstrating the ability of this method to recombine multiple, closely-spaced mutations.

4. Elimination Of The Parental Double Strands From Heteroduplex Preparations.

Annealing of substrates in vector-free form offers size-advantages relative to annealing of substrates as components of vectors, but does not allow selection for heteroduplexes relative to homoduplexes simply by transformation into host. Asymmetric PCR reactions with only one primer for each parent seeded with appropriate amount of previously amplified and purified gene fragment were run for 100 cycles, ensuring a 100-fold excess of one strand over another. Products of these asymmetrical reactions were mixed and annealed together producing only a minor amount of nonrecombinant duplexes. The last column in Table 3 shows the recombination efficiency obtained from these enriched heteroduplexes. Comparison of the first three

columns with the fourth one demonstrates the improvement achieved by asymmetric synthesis of the parental strands.

While the foregoing invention has been described in some detail for purposes of clarity and understanding, it will be clear to one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention. All publications and patent documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication or patent document were so individually denoted.

References

1. Shao, Z. and Arnold, F. H. 1996. Engineering new functions and altering existing functions. Curr. Opin. Struct. Biol. 6:513-518.

2. Kuchner, O and Arnold, F. H. 1997. Directed evolution of enzyme catalysts. Trends in Biotechnol. 15:523-530.

3. Abelson, J. N. (ed.) 1996. Combinatorial chemistry. Methods in Enzymol. 267, Academic Press, Inc. San Diego.

4. Joyce, G. F. 1992. Directed molecular evolution. Scientific American 267:90-97.

5. Stemmer, W. P. C. 1994a. Rapid evolution of a protein in vitro by DNA shuffling. Nature 370:389-391.

6. Stemmer, W. P. C. 1994b. DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. Proc. Nati. Acad. Sci. USA 91:10747-10751.

7. Moore, J. C. and Arnold, F. H. 1996. Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. Nature Biotech. 14:458-467.

8. Holland, J. H. 1975. Adaptation in natural and artificial systems. The University Press, Ann Arbor.

9. Goldberg, D. E. 1989. Genetic algorithms in search, optimization and machine learning. Addison-Wesley. Reading.

10. Eigen, M. 1971. Self-organization of matter and the evolution of biological macromolecules. Naturwissenschaften 58:465-523.

11. Rechenberg, L. 1973. Evolutions strategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution. Fronimann-Holzboog, Stuttgart.

12. Brady, R. M. 1985. Optimization strategies gleaned from biological evolution. Nature 317:804-806.

13. Muhlenbein, H. 1991. The parallel genetic algorithm as function optimizer. Parallel Computing 17:619-632.

14. Pal. K. F. 1993. Genetic algorithms for the traveling salesman problem-based on a heuristic crossover operation. Bio. Cybem. 69:539-546.

15. Pal. K. F. 1995. Genetic algorithm with local optimization. Bio.Cybem. 73:335-341.

16. Cami. B., P. Chambon. P. Kourilsky. 1984. Correction of complex heteroduplexes made of mouse H-2 gene sequences in *E. coli* K-12. Proc. Natl. Acad. Sci. USA 81:503-507.

17. Westmoreland. J. G. Porter, M. Radman and M. A. Resnick. 1997. Highly mismatched molecules resembling recombination intermediates efficiently transform mismatch repair proficient *E. coli*. Genetics 145:29-38.

18. Kramer. B., W. Kramer and H.-J. Fritz. 1984. Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch-repair system of *E. coli*. Cell 38:879-887.

19. Lu, A.-L., S. Clark and P. Modrich. 1983. Methyl-affected repair of DNA base pair mismatches in vitro. Proc. Natl. Acad. Sci. USA 80:4639-4643.

20. Carraway, M. and Marinus, M. G.. 1993. Repair of heteroduplex DNA molecules with multibase loops in *Escherichia coli*. J Bacteriol. 175:3972-3980.

21. Cooper, D. L., Lahue. R. S. and Modrich, P. 1993. Methyl-directed mismatch repair is bi-directional. J. Biol. Chem. 268:11823-11829.

22. Au, K. G., Welsh. K. and Modrich, P. 1992. Initiation of methyl-directed mismatch repair. J. Biol. Chem. 267:12142-12148.

23. Meselson, M. 1988. Methyl-directed repair of DNA mismatches, p. 91-113. In K. B. Low (ed.), Recombination of the Genetic Material. Academic Press, Inc., San Diego, Calif.

24. Fishel, R. A., Siegel, E. C. and Kolodner, R. 1986. Gene conversion in *Escherichia coli*. Resolution of heteroallelic mismatched nucleotides by co-repair. J. Mol. Biol. 188:147-157.

25. Pukkila, P. J., J. Peterson. G. Herman, P. Modrich. and M. Meselson. 1983. Effects of high levels of DNA adenine methylation on methyl-directed mismatch repair in *Escherichia coli*. Genetics 104:571-582.

26. Radman, M., R. E. Wagner, B. W. Glickman, and M. Meselson. 1980. DNA methylation, mismatch correction and genetic stability, p. 121-130. In M. Alacevic (ed.) Process in Environmental Mutagenesis. Elsevier/North-Holland Biochemical Press, Amsterdam.

27. Sambrook, J., Fritsch. E. F. and Maniatis, T. 1989. Molecular cloning: A Laboratory Manual. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

28. Allen, D. J., Makhov. A., Grilley, M...Taylor. J., Thresher. R., Modrich. P. and Griffith. J.D. MutS mediates heteroduplex loop formation by a translocation mechanism. 1997. EMBO J. 16: 4467-4476.

30. Tsai-Wu. J. J. and Lu. A. L. 1994. *Escherichia coli mutY*-dependent mismatch repair involves DNA polymerase I and a short repair tract. Mol. Gen. Genet. 244:444-450.

31. Worth, L. Jr., Clark, S., Radman. M. and Modrich, P. 1994. Mismatch repair proteins MutS and MutL inhibit RecA-catalyzed strand transfer between diverged DNAs. Proc. Natl. Acad. Sci. USA 91:3238-3241.

32. Fox, M. S., Radicella, J. P. and Yamamoto. K. 1994. Some features of base pair mismatch repair and its role in the formation of genetic recombinants. Experientia 50:253-260.

33. Radicella, J. P., Clark, E. A., Chen, S. and Fox, M.S. 1993. Patch length of localized repair events: role of DNA polymerase I in mutY-dependent mismatch repair. J. Bacteriol. 175: 7732-7736.

34. Kraczkiewicz-Dowjat, A. and Fishel, R. 1990. RecB-recC-dependent processing of heteroduplex DNA stimulates recombination of an adjacent gene in *Escherichia coli*. J. Bacteriol. 172:172-178.

35. Radman. M. 1989. Mismatch repair and the fidelity of genetic recombination. Genome 31: 68-73.

36. Raposa, S. and Fox, M. S. 1987. Some features of base pair mismatch and heterology repair in *Escherichia coli*. Genetics 117:381-390.

37. Jones, M., Wagner, R. and Radman, M. 1987. Mismatch repair and recombination in *E. coli*. Cell 50:621-626.

38. Langle-Rouault, F., Maenhaut-Michel, G. and Radman, M. 1987. GATC sequences, DNA nicks and the MutH function in *Escherichia coli* mismatch repair. EMBO J. 6:1121-1127

39. Glazer, P. M., Sarkar, S. N., Chisholm, G. E. and Summers. W. C. 1987. DNA mismatch repair detected in human cell extracts. Mol. Cell. Biol. 7:218-224

40. Laengle-Rouault, F., Maenhaut-Michel, G. and Radman M. 1986. GATC sequence and mismatch repair in *Escherichia coli*. EMBO J. 5:2009-2013.

41. Bauer, J., Krammer, G. and Knippers. R. 1981. Asymmetric repair of bacteriophage T7 heteroduplex DNA. Mol. Gen. Genet. 181:541-547.

42. Wildenberg, J. and Meselson. M. 1975. Mismatch repair in heteroduplex DNA. Proc. Natl. Acad. Sci. USA 72:2202-2206.

43. Kirkpatrick. D. T. and Petes. T. D. 1997. Repair of DNA loops involves DNA-mismatch and nucleotide-excision repair proteins. Nature 387: 929-31.

44. Leung, W., Malkova. A. and Haber. J. E. 1997. Gene targeting by linear duplex DNA frequently occurs by assimilation of a single strand that is subject to preferential mismatch correction. Proc. Natl. Acad. Sci. USA 94: 6851-6856.

45. Hunter, N. and Borts, R. H. 1997. Mlh1 is unique among mismatch repair proteins in its ability to promote crossing-over during meiosis. Genes Dev. 11:0890-9369.

46. Alani, E., Lee, S.. Kane. M. F.. Griffith, J. and Kolodner, R. D. 1997. *Saccharomyces cerevisiae* MSH2. a mispaired base recognition protein. also recognizes Holliday junctions in DNA. J. Mol. Biol. 265:289-301.

47. Varlet, I., Canard, B.. Brooks. P., Cerovic. G. and Radman. M. 1996. Mismatch repair in *Xenopus* egg extracts: DNA strand breaks act as signals rather than excision points. Proc. Natl. Acad. Sci. USA 93:10156-10161.

48. Nicolas, A. and Petes, T. D. 1994. Polarity of meiotic gene conversion in fungi: contrasting views. Experientia 50:242-52.

49. Bishop, D. K., J. Andersen, and R. D. Kolodner. 1989. Specificity of mismatch repair following transformation of *Saccharomyces cerevisiae* with heteroduplex plasmid DNA. Proc. Natl. Acad. Sci. USA 86:3713-3717.

50. Kramer, B., W. Kramer, M. S. Williamson. and S. Fogel. 1989. Heteroduplex DNA correction in *Saccharomyces cerevisiae* is mismatch specific and requires functional PMS genes Mol. Cell. Biol. 9:4432-4440.

51. Baynton, K., Bresson-Roy, A. and Fuchs, R. P. 1998. Analysis of damage tolerance pathways in *Saccharomyces cerevisiae*: a requirement for Rev3 DNA polymerase in translation synthesis. Mol. Cell. Biol. 18: 960-966.

52. Alani, E., Reenan, R. A. and Kolodner, R. D. 1994. Interaction between mismatch repair and genetic recombination in *Saccharomyces cerevisiae*. Genetics 137:19-39.

54. Bishop, D. K., Williamson. M. S., Fogel, S. and Kolodner. R. D. 1987.The role of heteroduplex correction in gene conversion in *Saccharomyces cerevisiae*. Nature 328:362-364.

55. Bishop, D. K. and Kolodner. R. D. 1986. Repair of heteroduplex plasmid DNA after transformation into *Saccharomyces cerevisiae*. Mol. Cell Biol. 6:3401-3409.

56. White. J. H., Lusnak. K. and Fogel, S. 1985. Mismatch-specific post-meiotic segregation frequency in yeast suggests a heteroduplex recombination intermediate. Nature 315: 350-352.

57. Abastado. J.-P., B. Cami. T. H. Dinh. J. Igoler and P. Kourilsky. 1984. Processing of complex heteroduplexes in *E. coli* and Cos-1 monkey cells. Proc. Natl. Acad. Sci. USA 81:5792-5796.

58. Brown. T. C. and J. Jiricny. 1987. A specific mismatch repair event protects mammalian cells from loss of 5-methylcytosine. Cell 50:945-950.

59. Sibghat-Ullah. and R-S. Day. 1993. DNA-substrate sequence specificity of human G:T mismatch repair activity. Nucleic Acids Res. 21:1281-1287.

60. Miller, E. M., Hough, H. L., Cho, J. W. and Nickoloff, J. A. 1997. Mismatch repair by efficient nick-directed, and less efficient mismatch-specific, mechanisms in homologous recombination intermediates in Chinese hamster ovary cells. Genetics 147: 743-753.

61. Deng, W. P. and Nickoloff, J. A. 1994. Mismatch repair of heteroduplex DNA intermediates of extrachromosomal recombination in mammalian cells. Mol. Cell Biol. 14:400-406.

62. Thomas, D. C., Roberts, J. D. and Kunkel, T. A. 1991. Heteroduplex repair in extracts of human HeLa cells. J. Biol. Chem. 266:3744-51.

63. Folger, K. R., Thomas, K. and Capecchi, M. R. 1985. Efficient correction of mismatched bases in plasmid heteroduplexes injected into cultured mammalian cell nuclei. Mol. Cell. Biol. 5:70-74.

64. Fang, W., Wu, J. Y. and Su, M. J. 1997. Methyl-directed repair of mismatched small heterologous sequences in cell extracts from *Escherichia coli*. J. Biol. Chem. 272: 22714-22720.

65. Smith, J. and Modrich, P. 1997. Removal of polymerase-produced mutant sequences from PCR products. Proc. Natl. Acad. Sci. U S A 94: 6847-50.

66. Su, S. S., Grilley, M., Thresher, R., Griffith, J. and Modrich, P. 1989. Gap formation is associated with methyl-directed mismatch correction under conditions of restricted DNA synthesis. Genome 31:104-11.

67. Muster-Nassal, C. and Kolodner, R. 1986. Mismatch correction catalyzed by cell-free extracts of *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA 83:7618-7622.

68. Macnab, R.M. 1992. Genetic and biogenesis of bacterial flagella. Annul Rev. Genet. 26:131-158.

69. Wilson, D. R. and Beveridge, T. J. 1993. Bacterial flagellar filaments and their component flagellins. Can. J. Microbiol. 39:451-472.

70. Schmitt, R., Raskal, A. and Mayer, F. 1974. Plain and complex flagella of *Pseudomonas rhodos*: analysis of fine structure and composition. J. Bacteriol. 117:844-857.

71. Gotz, R., Limmer, N., Ober, K. and Schmitt, R. 1982. Motility and chemotaxis in two strains of *Rhizobium* with complex flagella. J. Gen. Microbiol. 128:789-798.

72. Schmitt, R., Bambergerl, Acker G. and Mayer, F. 1974. Fine structure analysis of the complex flagella of *Rhizobium lupini* H13-3. Arch. Microbiol. 100:145-162.

73. Trachtenberg, S., DeRosier, D. J. and Macnab, R. M. 1987. Three-dimensional structure of the complex flagellar filament of *Rhizobium lupini* and its relation to the structure of the plain filaments. J. Mol. Biol. 195:603-620,

74. Götz, R. and Schmitt, R. 1987. *Rhizobium meliloti* swims by unidirectional intermittent rotation of right-handed flagellar helices. J. Bacteriol. 169:3146-3150.

75. Lotz, W., Acker, G. and Schmitt. R. 1977. Bacteriophage 7-7-1 adsorbs to the complex flagella of *Rhizobium lupini* H13-3. J. Gen. Virol. 34:9-17.

76. Krupski, G.,Gotz, F., Ober, K., Pleicr, E. and Schmitt, R. 1985. Structure of complex flagellar filaments in *Rhizobium meliloti*. J. Bacteriol. 162:361-366.

77. Maruyama, M., Lodderstaedt, G. and Schmitt, R. 1978. Purification and. biochemical properties of complex flagella isolated from *Rhizobium lupini* H13-3. Biochem. Biophys. Acta 535:110-124.

78. Trachtenberg, S., DeRosier, D. J., Aizawa, S.-I. and Macnab, R. M. 1986. Pairwise perturbation of flagellin subunits. The structural basis for the differences between plain and complex bacterial flagellar filaments. J. Mol. Biol. 190:569-576.

79. Gordee, R. S., Zeckner, D. J., Ellis, L. F., Thakkar, A. L. and Howard, L. C. 1984. In vitro and in vivo anti-*Candida* activity and toxicity of LY121019. J. Antibiotics 37:1054-1065.

80. Debono, M., Willard, K. E., Kirst, H. A., Wind, J. A., Crouse, G. D., Tao, E. V., Vicenzi, J. T., Counter, F. T., Ott, J. L., Ose, E. E. and Omura, S. 1989. Synthesis of new analogs of echinocandin B by enzymatic deacylation and chemical reacylation of the echinocandin B peptide: synthesis of the antifungal agent cilofungin (LY121019). J. Antibiotics 42(3): 389-397.

81. Debono, M. and Gordee, R. S. 1994. Antibiotics that inhibit fungal cell-wall development. Annu. Rev. Microbiol. 48: 471-497.

82. Debono, M., Turner, W. W., Lagrandeur, L., Burkhardt, F. J., Nissen. J. S., Nichols, K. K., Rodriguez, M. J., Zweifel, M. J., Zeckner, D. J., Gordee, R. S., Tang. J. and Parr, T. R. 1995. Semisynthetic chemical modification of the antifungal lipopeptide echinocandin B (ECB): structure-activity studies of the lipophilic and geometric parameters of polyarylated acyl analogs of ECB. J. Med. Chem. 38(17): 3271-3281.

83. Yeh, W. K. 1997. Evolving enzyme technology for pharmaceutical applications: case studies. J. Ind. Microbiol. Biotechnol. 19(5-6): 334-343.

84. Boeck, L. D., Fukuda, D., Abbott. B. J. and M. Debono. 1989. Deacylation of echinocandin B by *Actinoplanes utahensis*. J. Antibiotics 42(3): 382-388

85. Arnold, F. H. 1998. Design by directed evolution. Accts. Chem. Res. 31:125-131.

86. Shao, Z., Callahan. M. and Arnold. F. H. 1998. Directed enzyme evolution of *Actinoplane utahensis* ECB deacylase in *Streptomyces lividans* for enhanced specific activity. Manuscript submitted.

87. Shao, Z., Zhao. H.. Giver. L. and Arnold, F. H. 1998. Random-priming in vitro recombination: an effective tool for directed evolution. Nucleic Acids Res. 26 (2): 681-683.

88. Zhao, H. and Arnold. FH, 1997. Functional and nonfunctional mutations distinguished by random recombination of homologous genes. Proc. Natl. Acad. Sci. USA 94:7997-8000.

89. Zhao, H., Giver, L., Shao, Z., Affholter, J.A., and Arnold, F.H. 1998. Molecular evolution by staggered extension process (StEP) in vitro recombination. Nat. Biotechnol. 16: 258-261.

90. Judo. M. S. B.. Wedel. A. B. and Wilson, C. 1998. Stimulation and suppression of PCR-mediated recombination. Nucleic Acids Res. 26: 1819-1825.

91. Okkels, J. S. 1997. Method for preparing polypeptide variants. PCT application WO 97/07205.

92. Gray, G. L. 1992. Hybrid prokaryotic polypeptides produced by in vivo homologous recombination. U.S. Patent 5,093,257.

93. Weber, H. and Weissmann, C. 1983. Formation of genes coding for hybrid proteins by recombination between related, cloned genes in *E. coli*. Nucl. Acids Res. 11:5661-5669.

94. Maryon, E. and Carroll. D. 1991. Characterization of recombination intermediates from DNA injected into Xenopus laevis oocytes: evidence for a nonconservative mechanism of homologous recombination. Mol. Cell. Biol. 11:3278-3287.

WHAT IS CLAIMED IS:

1. A method for evolving a polynucleotide toward acquisition of a desired property, comprising

(a) incubating a population of parental polynucleotide variants under conditions to generate annealed polynucleotides comprising heteroduplexes;

(b) exposing the heteroduplexes to a cellular DNA repair system to convert the heteroduplexes to parental polynucleotide variants or recombined polynucleotide variants;

(c) screening or selecting the recombined polynucleotide variants for the desired property.

2. The method of claim 1, wherein the heteroduplexes are exposed to the cellular DNA repair system in vitro.

3. The method of claim 2, wherein the cellular DNA repair system comprises cellular extracts.

4. The method of claim 1, further comprising introducing the heteroduplexes into cells, whereby the heteroduplexes are exposed to the DNA repair system of the cells in vivo.

5. The method of claim 4, wherein the annealed polynucleotides further comprise homoduplexes and the introducing step selects for transformed cells comprising the heteroduplexes relative to transformed cells comprising homoduplexes.

6. The method of claim 4, wherein a first polynucleotide variant is provided as a component of a first vector, and a second polynucleotide variant is provided as a component of a second vector, and the method further comprises converting the first and second vectors to linearized forms in which the first and second polynucleotide variants occur at opposite ends, whereby in the incubating step single-stranded forms of the first linearized vector reanneal with each other to form linear first vector, single-stranded forms of the second linearized vector reanneal with each other to form linear second vector, and single-stranded linearized forms of the first and second vectors anneal with each to form a circular heteroduplex bearing a nick in each strand, and the

introducing step selects for transformed cells comprising the circular heteroduplexes relative to the linear first and second vector.

7. The method of claim 6. wherein the first and second vectors are converted to linearized forms by PCR.

8. The method of claim 6, wherein the first and second vectors are converted to linearized forms by digestion with first and second restriction enzymes.

9. The method of claim 1, wherein the population of polynucleotide variants are provided in double stranded form, and the method further comprising converting the double stranded polynucleotides to single stranded polynucleotides before the annealing step.

10. The method of claim 1, wherein the converting step comprises:

conducting asymmetric amplification of the first and second double stranded polynucleotide variants to amplify a first strand of the first polynucleotide variant, and a second strand of the second polynucleotide variant, whereby the first and second strands anneal in the incubating step to form a heteroduplex.

11. The method of claim 10, wherein the first and second double-stranded polyncueltoides variants are provided in vector-free form. and the method further comprises incorporating the heteroduplex into a vector.

12. The method of claim 4 wherein the population of polynucleotides comprises first and second polynucleotides provided in double stranded form, and the method further comprises incorporating the first and second polynucleotides as components of first and second vectors, whereby the first and second polynucleotides occupy opposite ends of the first and second vectors. whereby in the incubating step single-stranded forms of the first linearized vector reanneal with each other to form linear first vector. single-stranded forms of the second linearized vector reanneal with each other to form linear second vector, and single-stranded linearized forms of the first and second vectors anneal with each to form a circular heteroduplex bearing a nick in each

strand. and the introducing step selects for transformed cells comprises the circular heteroduplexes relative to the linear first and second vector.

13. The method of claim 4, further comprising sealing nicks in the heteroduplexes to form covalently-closed circular heteroduplexes before the introducing step.

14. The method of claim 11, wherein the first and second polynucleotides are obained from chromosomal DNA..

15. The method of claim 1, further comprising repeating steps (a)-(c) whereby the incubating step in a subsequent cycle is performed on recombinant variants from a previous cycle.

16. The method of claim 1, wherein the polynucleotide variants encode a polypeptide.

17. The method of claim 1, wherein the population of polynucleotide variants comprises at least 20 variants.

18. The method of claim 1, wherein the population of polynucleotide variants are at least 10 kb in length.

19. The method of claim 1, wherein the population of polynucleotide variants comprises natural variants.

20. The method of claim 1, wherein the population of polynucleotides comprises variants generated by mutagenic PCR.

21. The method of claim 1, wherein the population of polynucleotide variants comrises variants generated by site directed mutagenesis.

22. The method of claim 1. wherein the cells are bacterial cells.

23. The method of claim 1. further comprising at least partially demethylating the population of variant polynucleotides.

24. The method of claim 23, whether the at least partially demethylating step is performed by PCR amplification of the population of variant polynucleotides.

25. The method of claim 23, wherein the at least partially demethylating step is performed by amplification of the population of variant polynucelotides in host cells.

26. The method of claim 25, wherein the host cells are defective in a gene encoding a methylase enzyme.

27. The method of claim 1, wherein the population of variant polynucleotide variants comprises at least 5 polynucleotides having at least 90% sequence identity with one another.

28. The method of claim 1, further comprising isolating a screened recombinant variant.

29. The method of claim 28, further comprising expressing a screened recombinant variant to produce a recombinant protein.

30. The method of claim 29 further comprising formulating the recombinant protein with a carrier to form a pharmaceutical composition.

31. The method of claim 1, wherein the polynucleotide variants encode enzymes selected from the group consisting of proteases, lipases, amylases. cutinases, cellulases, amylases. oxidases, peroxidases and phytases.

32. The method of claim 1, wherein the polynucleotide variants encode a polypeptide selected from the group consisting of insulin. ACTH. glucagon. somatostatin. somatotropin. thymosin. parathyroid hormone. pigmentary hormones. somatomedin. erthropoietin. luteinizing hormone. chorionic gonadotropin. hyperthalmic releasing factors. antidiuretic hormones. thyroid stimulating hormone. relaxin. interferon. thrombopoietin (TPO). and prolactin.

33. The method of claim 1, wherein the polynucleotide variants encode a plurality of enzymes forming a metabolic pathway.

34. The method of claim 1, wherein the polynucleotide variants are in concatemeric form.

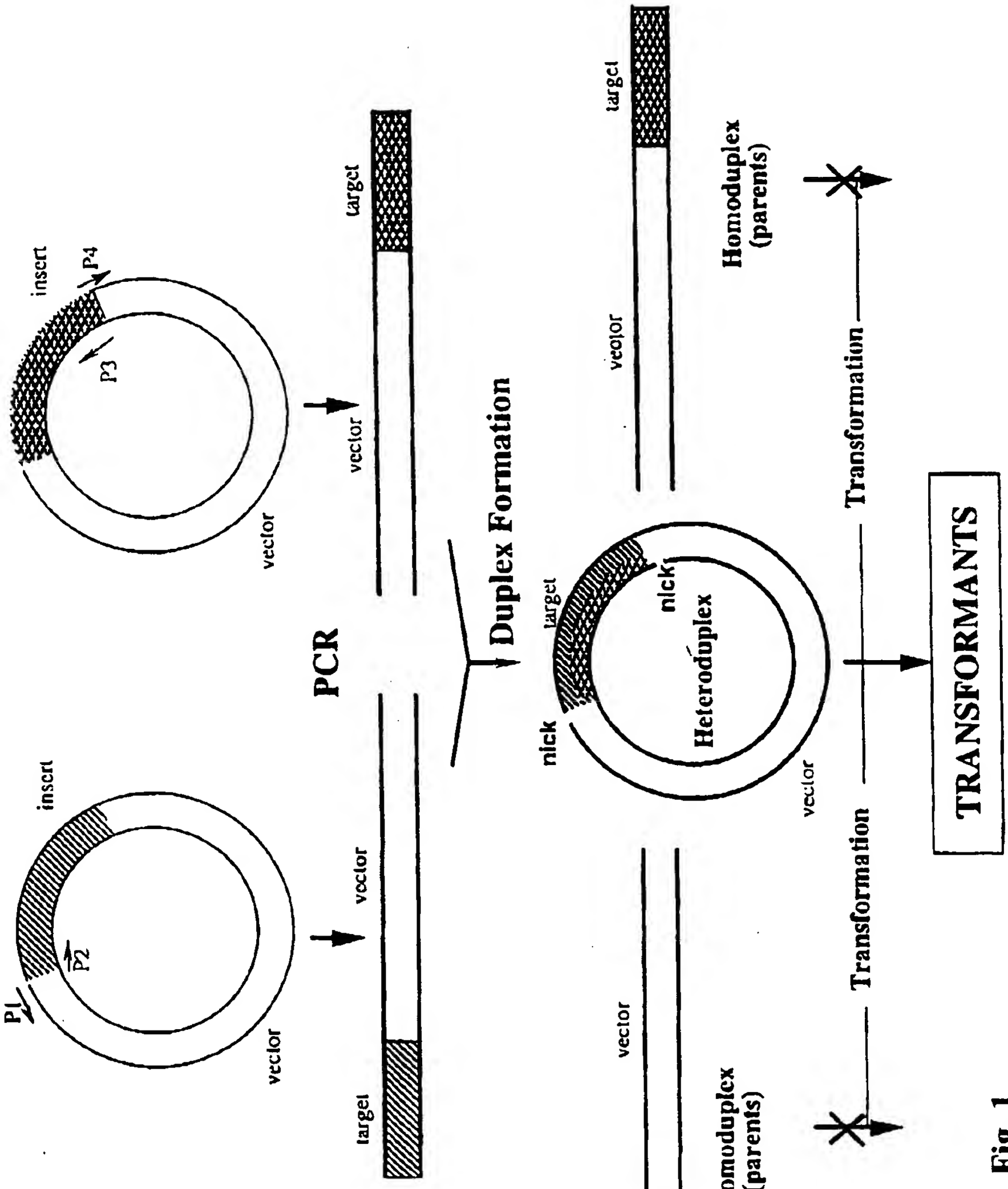

insert
P4
P3
vector
P1
P2
insert
vector
PCR
target
vector
Duplex Formation
nick
target
nick
Heteroduplex
vector
target
vector
Homoduplex
(parents)
Transformation
TRANSFORMANTS
target
vector
Homoduplex
(parents)
Transformation

Fig. 1

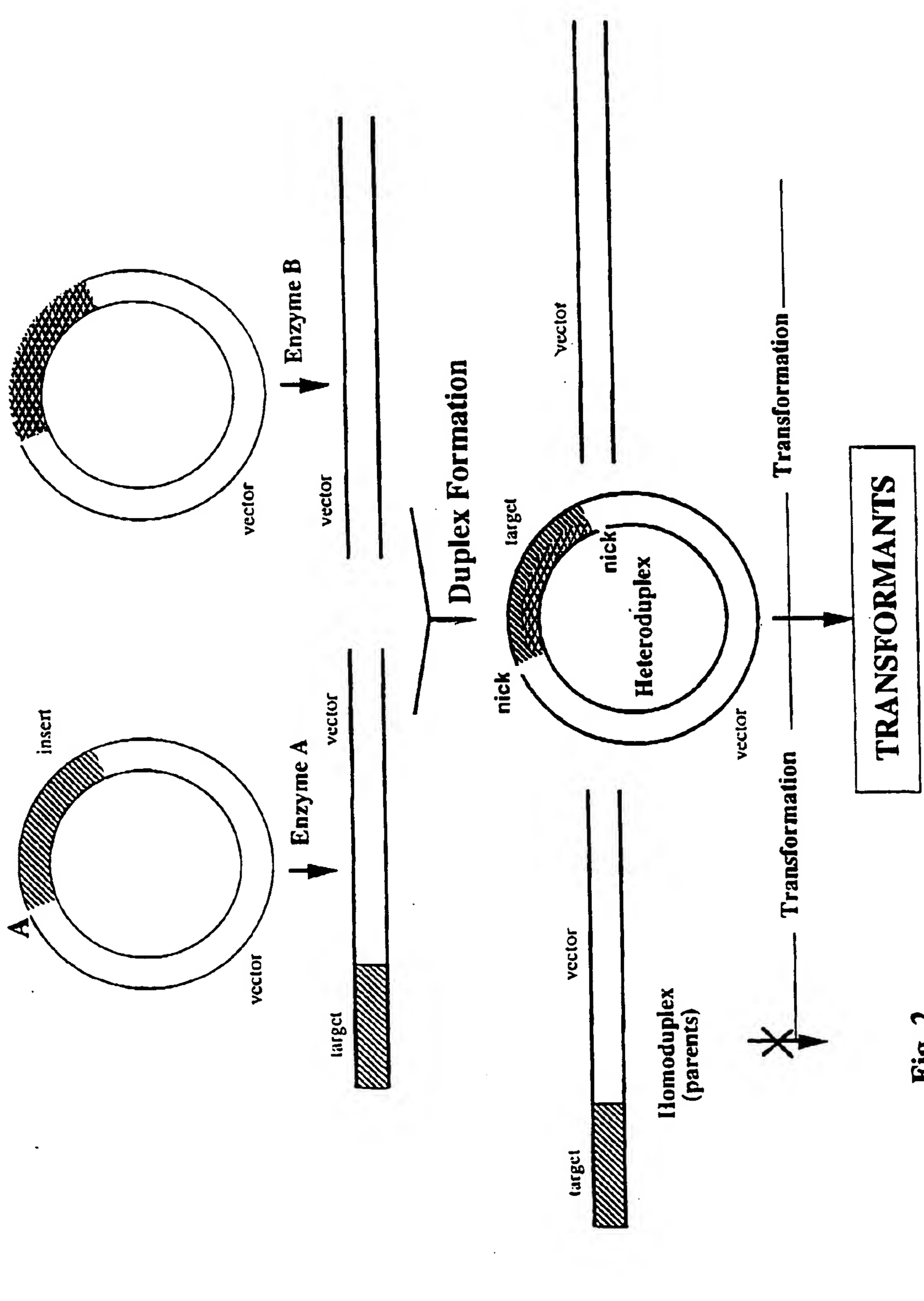

insert
A
vector
Enzyme A
vector
target
Enzyme B
vector
vector
Duplex Formation
target
nick
nick
Heterodiplex
vector
vector
Transformation
Transformation
Homoduplex
(parents)
target
vector
vector
TRANSFORMANTS

Fig. 2

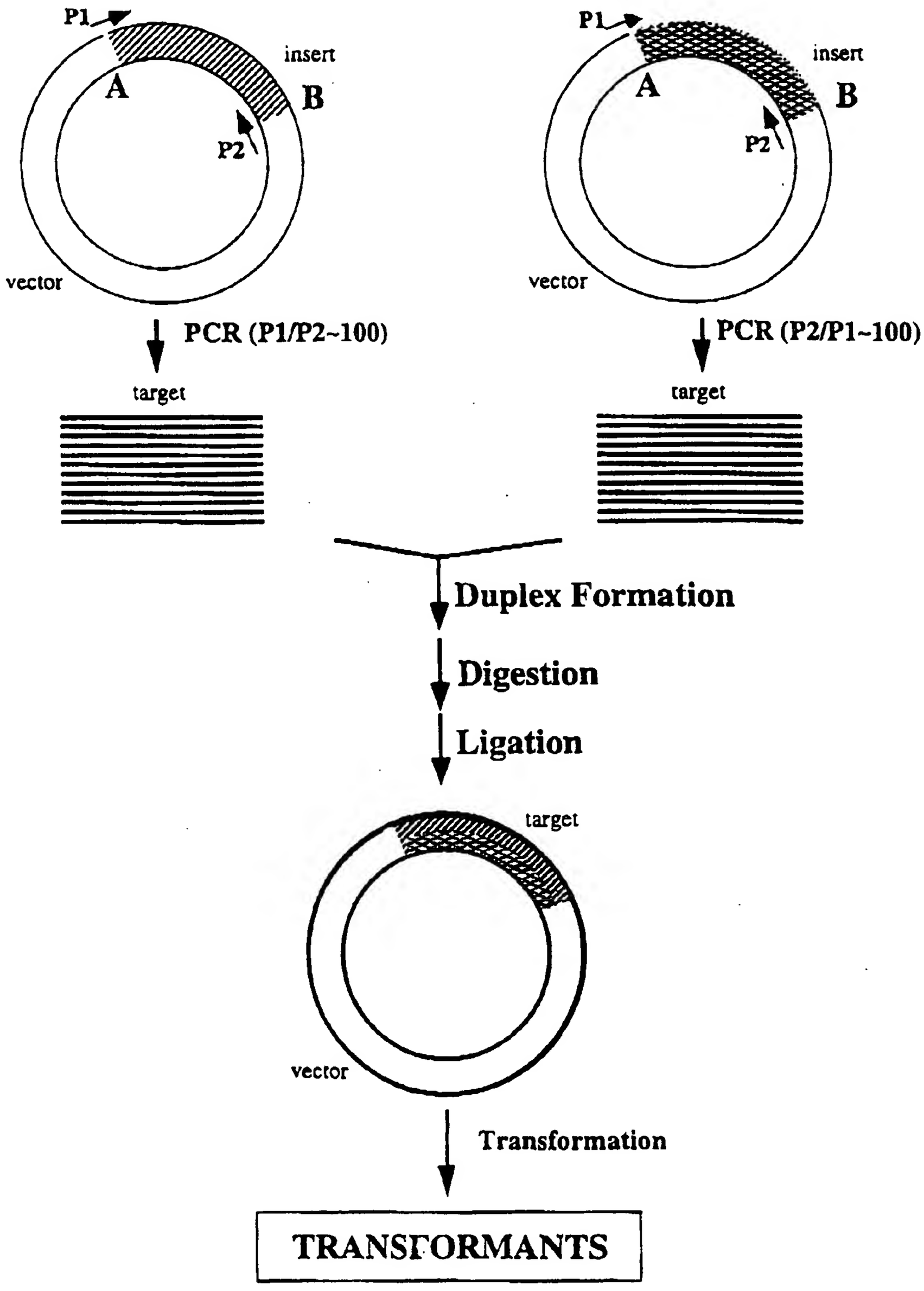

P1
insert
A
B
P2
vector
PCR (P1/P2~100)
target
P1
insert
A
B
P2
vector
PCR (P2/P1~100)
target
Duplex Formation
Digestion
Ligation
target
vector
Transformation
TRANSFORMANTS

Fig. 3

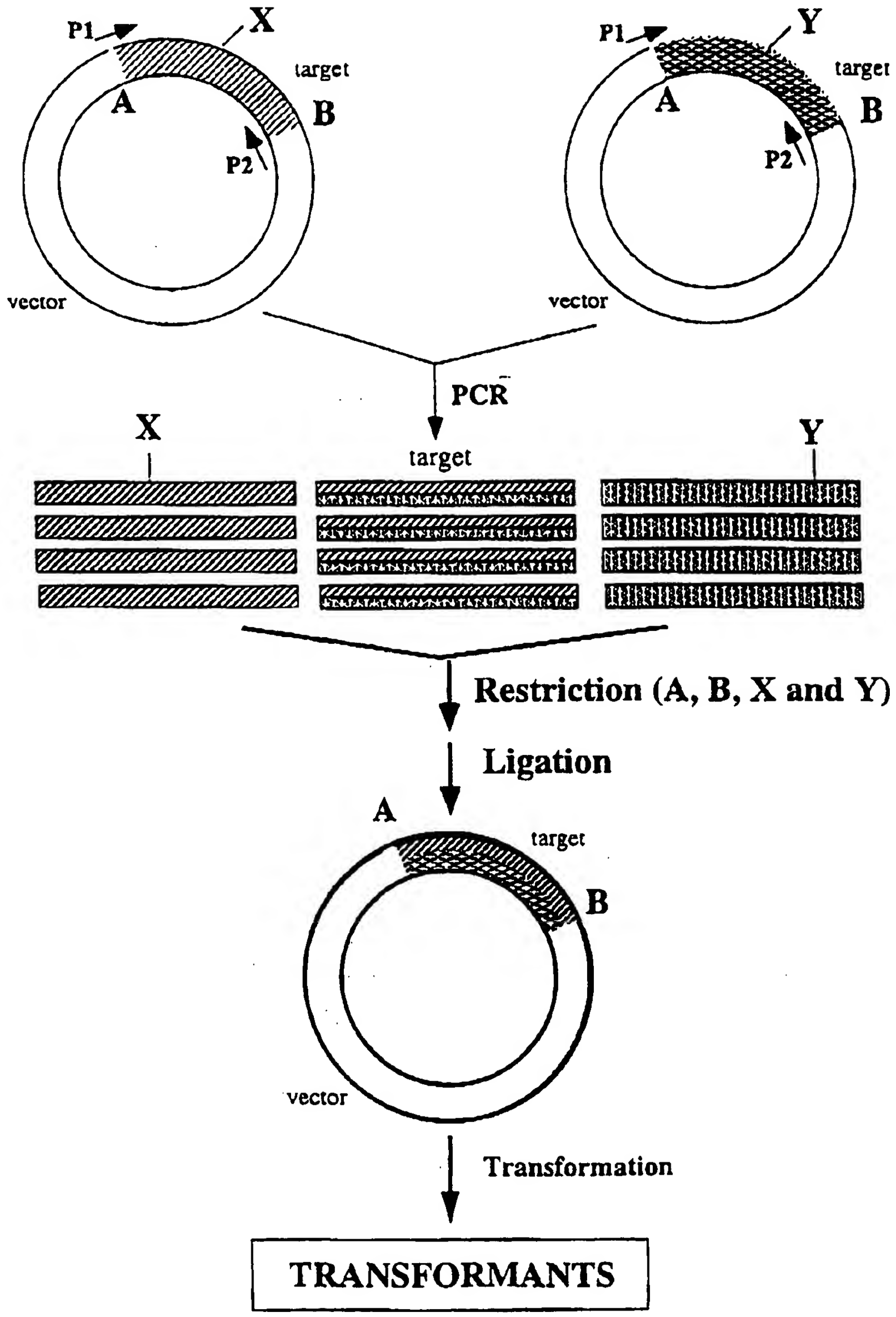

P1
X
target
A
B
P2
vector
P1
Y
target
A
B
P2
vector
PCR
X
target
Y
Restriction (A, B, X and Y)
Ligation
A
target
B
vector
Transformation
TRANSFORMANTS

Fig. 4

```
  1 MASVLTNINAMSALQTLRSISSNMEDTQSRISSGMRVGSASDNAAYWSIATTMRSDNASLSAVQDAIGLG  R.lupini FlaA
    |.|:||| .||.|| |||||||.|||||||||||:||||||||||||||||||||||..||||||||:|||
  1 MTSILTNNSAMAALSTLRSISSSMEDTQSRISSGLRVGSASDNAAYWSIATTMRSDNQALSAVQDALGLG  R.meliloti FlaA

 71 AAKVDTASAGMDAVIDVVKQIKNKLVTAQESSADKTKIQGEVKQLQEQLKGIVDSASFSGENWLKGDLST  R.lupini FlaA
    ||||||| .||:..|:|||:||.|||.| |.:.||.|||:|:.||.:||.:|.:.|||||||||.:|||.
 71 AAKVDTAYSGMESAIEVVKEIKAKLVAATEDGVDKAKIQEEITQLKDQLTSIAEAASFSGENWLQADLSG  R.meliloti FlaA

141 TT.TKSVVGSFVRE.GGTVSVKTIDYALNASKVLVDTRATGTKTGILDTAYTGLNANTVTVDINKGGVIT  R.lupini FlaA
    .. ||||||:|||: :|.||||.:||.||...||.|   |...|||||..| .:....||:.:| .|...
141 GPVTKSVVGGFVRDSSGAVSVKKVDYSLNTDTVLFD...TTGNTGILDKVY.NVSQASVTLPVNVNGTTS  R.meliloti FlaA

209 QASVRAYSTDEMLSLGAKVDGANSNVAVGGGSASSRSTAAGLRVASTLRPPSPEQHQSLASLPPLTPPLK  R.lupini FlaA
    : .| ||..|:::. :|..||. .||:.|: .:.  ...::: | .: :.:..
207 EYTVGAYNVDDLIDASATFDGDYANVGAGALAGDYVKVQGSWVKAVDVAATGQE.............  R.meliloti FlaA

279 LVLQLLPVTPSSSTKPTAAP.VQVNLTQSVLTMDVS.SMSSTDVGSYLTGVEKALTSLTSAGAELGSIKQ  R.lupini FlaA
    :|.:  ... : .|..|:|| ..|... |: |:|:. .  ..:::..:.||:.|||.:|||:|.|||||.
277 VVYDDGTTKWGVDTTVTGAPATNVAAPASLATIDITIAAQAGNLDALIAGVDEALTDMTSAAASLGSISS  R.meliloti FlaA

347 RIDLQVDFASKLGDALAKGIGRLVDADMNEESTKLKALQTQQQLAIQSLSIANSDSQNILSLFR 410    R.lupini FlaA
    ||||| ||..||:|.:..|:|||||||||||||:||||||||||||.|||||||||||:|||||
331 RIDLQSDFVNKLSDSIDSGVGRLVDADMNEESTRLKALQTQQQLAIQALSIANSDSQNVLSLFR 394    R.meliloti FlaA
```

Fig. 5

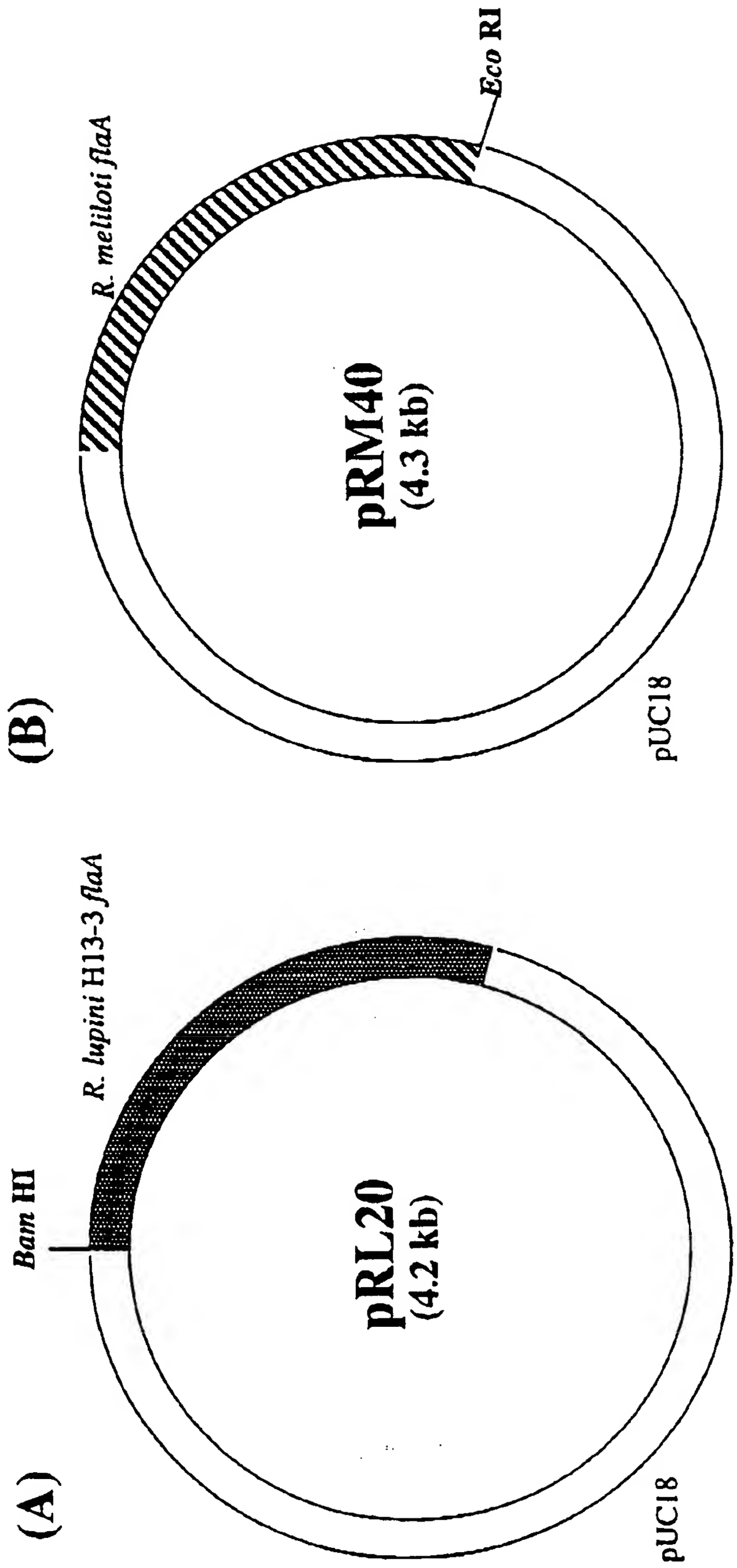

(A)
Bam HI
R. lupini H13-3 flaA
pRL20
(4.2 kb)
pUC18
(B)
Eco RI
R. meliloti flaA
pRM40
(4.3 kb)
pUC18

Fig. 6A & 6B

(a)

```
SCS01   ATGGCAAGCGTTCTCACAAACATTAACGCAATGTCTGCTCTTCAGACGCTGCGTTCGATTTCTTCCAACATGGAAGACACCCAGAGCCGTATTTCCAGCG
SCS01   GCATGCGCGTTGGTTCGGCTTCCGACAACGCCGCTTATTGGTCTATCGCGACCACCATGCGCTCGGACAATGCCTCGCTTTCCGCTGTTCAGGATGCAAT
SCS01   TGGCCTCGGTGCCGCCAAGGTCGATACCGCTTCGGCGGGTATGGATGCGGTTATCGATGTTGTAAAGCAGATCAAGAACAAACTGGTCACTGCCACCGAA
SCS01   GACGGCGTCGACAAGGCCAAGATCCAAGAAGAAATCACTCAGCTCAAGGACCAGCTGACGAGCATCGCCGACGCGGCTTCCTTCTCCGGTGAAAACTGGC
SCS01   TCAAGGGCGATCTTTCCACGACGAC.....AACCAAATCAGTGGTTGGCTCCTTCGTTCGTGA...AGGCGGTACCGTATCGGTCAAGACCATCGATTAC
SCS01   GCTCTGAATGCTTCCAAGGTTCTGGTGGATACCCGCGCAACGGGCACCAAGACCGGCATTCTGGACAAGGTCTACAACGTCTCGCAGGCAAGCGTCACGC
SCS01   TGACGG.....TCAACACCAACGGCGTCGAATCCCAGGCCTCCGTCCGCGCCTATtCGCTGGA.....GTCCCTCACCGAAGCCGGTGCGGAGTT
SCS01   CCAGGGCAACTATGCTCTTCAGGGCGGTAACAGCTACGTCAAGGTCGAAAACGTCTGGGT.....TCGA.GCTGAG.....ACCGCATCA
SCS01   ACACCAGTCGCTGGCAAGTTTGCCGCCGCTTACACCGCCGCTGAAGCTGGTACTGCAGCTGCTGCCGGTGACgCCATCATCGTCGACGAAACCAACAGCG
SCS01   GCGCCGGTGCAGGTAAACCTCACCCAGTCGgTCCTGACCATGGATGTCAGCTCGATGAGCTCGACGGATGTCGGCAGCTACCTCACGGGCGTGGAAAAGG
SCS01   CTCTCACCAGCCTGACGAGCGCTGGCGCTGAACTCGGCTCTATCAAACAGCGCATCGATCTGCAGGTTGATTTTGCTTCCAAgctGGGCGACGCTCTCGC
SCS01   AAAAGGTATTGGCCGTCTCGTTGATGCTGACATGAATGAAGAGTCCACTAAGCTTAAGGCTCTTCAGACGCAGCAGCAGCTGGCTATCCAGTCGCTCTCC
SCS01   ATCGCAAACAGCGACTCGCAGAACATTCTGTCGCTGTTCCGTTAA
```

(b)

```
SCS02   ATGACGAGCATTCTCACCAACAACTCCGCAATGGCCGCGCTTTCCGGAGTGCGCTCGATCTCTTCCAGCATGGAAGACACGCAGAGCCGCATCTCCTCCG
SCS02   GCCTTCGCGTCGGTTCGGCCTCCGACAACGCCGCCTACTGGTCGATTGCGACCACCATGCGCTCCGACAACCAGGCCCTTTCGGCCGTCCAGGACGCCCT
SCS02   CGGCCTCGGCGCCGCCAAGGTTGATACCGCCTATTCCGGTATGGAATCGGCGATCGAAGTCGTTAAGGAAATCAAGAACAAACTGGTCACTGCTCAGGAA
SCS02   TCTTCTGCCGACAaAACGaAGATTCAGGGCGAAGTCAAGCAGCTTCAGGAGCAGTTGAAGGGCaTCGTTGATTCCGCTTCCTTCTCCGGTGAGAACTGGC
SCS02   TGCAGGCGGACC..TCAGCGGCGGCGCCGTCACCAAGAGCGTCGTCGGCTCGTTCGTCCGTGACGGAAGCGGTTCCGTAGCCGTCAAGAAGGTCGATTAC
SCS02   GCTCTGAATGCTTCCAAGGTTCTGGTGGATACCCGCGCAACGGGCACCAAGACCGGCATTCTCGATACTGCTTATACCGGCCTTAACGCGA...ACACGG
SCS02   TGACGGTTGATATCAACAAGGGCGGCGTGATCACCCAGGCCTCCGTCCGCGCCTATtCCACgGACGAAATGCTCTCCCTCGGCGCAAAGGTCGATGGCGC
SCS02   AAACAGCAACGTTGCTGTTGGCGGCGG.CTCCGCTTCGTcAAGGTCGACGGCAGCTGGGTTAAGGGTAGCGTCGACqCTGCGGCCTCCATCACCGCATCA
SCS02   ACCGGCGCCACCGGTCAAGAAATCGCCGC...CACCACGACGGCAGCTGGTACCATCACTGCAGACAGCTGGGTCGTCGATGTCGGCAACGCTCCTGCCG
SCS02   CCAACGTTTCGG.......CCGGCCAGTCGGTCGCGAACATCAACATCGTCGGAATGGGCTCGACGGATGTCGGCAGCTACCTCACGGGCGTGGAAAAGG
SCS02   CTCTCACCAGCATGACCAGCGCTGCCGCCTCGCTCGGCTCCATCTCCTCGCGCATCGACCTGCAGAGCGAATTCGTCAACAAGCTCTCGGACTCGATCGA
SCS02   GTCGGGCGTCGGCCGTCTCGTCGACGCGGACATGAACGAGGAGTCGACCCGCCTCAAGGCCCTGCAGACCCAGCAGCAGCTCGCCATCCAGGCCCTGTCG
SCS02   ATCGCCAACTCGGACTCGCAGAACGTCCTGTCGCTCTTCCGCTAA
```

Fig. 7A & 7B

(c)

```
ES01    ATGACGAGCATTCTCACCAACAACTCCGCAATGGCCGCGCTTTCCGGAGTGCGCTCGATCTCTTCCAGCATGGAAGACACGCAGAGCCGCATCTCCTCCG
ES01    GCCTTCGCGTCGGTTCGGCCTCCGACAACGCCGCCTACTGGTCGATTGCGACCACCATGCGCTCCGACAACCAGGCCCTTTCGGCCGTCCAGGACGCCCT
ES01    CGGCCTCGGCGCCGCCAAGGTTGATACCGCCTATTCCGGTATGGAATCGGCGATCGAAGTCGTTAAGGAAATCAAGGCCAAGCTCGTAGCTGCCACCGAA
ES01    GACGGCGTCGACAAGGCCAAGATCCAAGAAGAAATCACTCAGCTCAAGGACCAGCTGACGAGCATCGCCGACGCGGCTTCCTTCTCCGGTGAGAACTGGC
ES01    TGCAGGCGGACC..TCAGCGGCGGCGCCGTCACCAAGAGCGTCGTCGGCTCGTTCGTCCGTGACGGAAGCGGTTCCGTAGCCGTCAAGACCATCGATTAC
ES01    GCTCTGAATGCTTCCAAGGTTCTGGTGGATACCCGCGACACGGTCGGCGATACCGGCATTCTGGACAAGGTCTACAACGTCTCGCAGGCAAGCGTCACGC
ES01    TGACGG......TCAACACCAACGGCGTCGAATCGCAGCATACGGTTGCTGCCTATTCGCTGGA.........GTCCCTCACCGAAGCCGGTGCGGAGTT
ES01    CCAGGGCAACTATGCTCTTCAGGGCGGTAACAGCTACGTCAAGGTCGACGGCAGCTGGGTTAAGGGTAGCGTCGACGCTGCGGCCTCCATCACCGCATCA
ES01    ACACCAGTCGCTGGCAAGTTTGCCGCCGCTTACACCGCCGCTGAAGCTGGTACTGCAGCTGCTGCCGGTGACGCCATCATCGTCGACGAAACCAACAGCG
ES01    GCGCCGGTGCAGGTAAACCTCACCCAGTCGCGTCCTGACCATGGATGTCAGCTCGATGAGCTCGACGGATGTCGGCAGCTACCTCACGGGCGTGGAAAAGG
ES01    CTCTCACCAGCCTGACGAGCGCTGGCGCTGAACTCGGCTCCATCTCCTCGCGCATCGACCTGCAGAGCGAATTCGTCAACAAGCTCTCGGACTCGATCGA
ES01    GTCGGGCGTCGGCCGTCTCGTCGACGCGGACATGAACGAGGAGTCGACCCGCCTCAAGGCCCTGCAGACCCAGCAGCAGCTCGCCATCCAGGCCCTGTCG
ES01    ATCGCCAACTCGGACTCGCAGAACGTCCTGTCGCTCTTCCGCTAA
```

(d)

```
ES02    ATGACGAGCATTCTCACCAACAACTCCGCAATGGCCGCGCTTTCCGGAGTGCGCTCGATCTCTTCCAGCATGGAAGACACGCAGAGCCGCATCTCCTCCG
ES02    GCCTTCGCGTCGGTTCGGCCTCCGACAACGCCGCCTACTGGTCGATTGCGACCACCATGCGCTCCGACAACCAGGCCCTTTCGGCCGTCCAGGACGCCCT
ES02    CGGCCTCGGCGCCGCCAAGGTTGATACCGCCTATTCCGGTATGGAATCGGCGATCGAAGTCGTTAAGGAAATCAAGGCCAAGCTCGTAGCTGCCACCGAA
ES02    GACGGCGTCGACAAGGCCAAGATCCAAGAAGAAATCACTCAGCTCAAGGACCAGCTGACGAGCATCGCCGACGCGGCTTCCTTCTCCGGTGAGAACTGGC
ES02    TGCAGGCGGACC..TCAGCGGCGGCGCCGTCACCAAGAGCGTCGTCGGCTCGTTCGTCCGTGACGGAAGCGGTTCCGTAGCCGTCAAGACCATCGATTAC
ES02    GCTCTGAATGCTTCCAAGGTTCTGGTGGATACCCGCGACAACGGGCACCAAGACCGGCATTCTCGATACTGCTTATACCGGCCTTAACGCGA...ACACGG
ES02    TGACGGTTGATATCAACAAGGGCGGCGTGATCACCCAGGCCTCCGTCCGCGCCTATTCCACGGACGAAATGCTCTCCCTCACCGAAGCCGGTGCGGAGTT
ES02    CCAGGGCAACTATGCTCTTCAGGGCGGTAACAGCTACGTCAAGGTCGAAAACGTCTGGGT..........TCGA.GCTGAG.........ACCGCTGCA
ES02    ACCGGCGCCACCGGTCAAGAAATCGCCGC...CACCACGACGGCAGCTGGTACCATCACTGCAGACAGCTGGGTCGTCGATGTCGGCAACGCTCCTGCCG
ES02    CCAACGTTTCGG.......CCGGCCAGTCGGTCGCGAACATCAACATCGTCGGAATGGGTGCAGCTGCGCTCGATGCCCTGATCAGCGGTGTCGACGCCG
ES02    CTTTGACAGACATGACCAGCGCTGCCGCCTCGCTCGGCTCCATCTCCTCGCGCATCGACCTGCAGAGCGAATTCGTCAACAAGCTCTCGGACTCGATCGA
ES02    GTCGGGCGTCGGCCGTCTCGTCGACGCGGACATGAACGAGGAGTCGACCCGCCTCAAGGCCCTGCAGACCCAGCAGCAGCTCGCCATCCAGGCCCTGTCG
ES02    ATCGCCAACTCGGACTCGCAGAACGTCCTGTCGCTCTTCCGCTAA
```

Fig. 7C & 7D

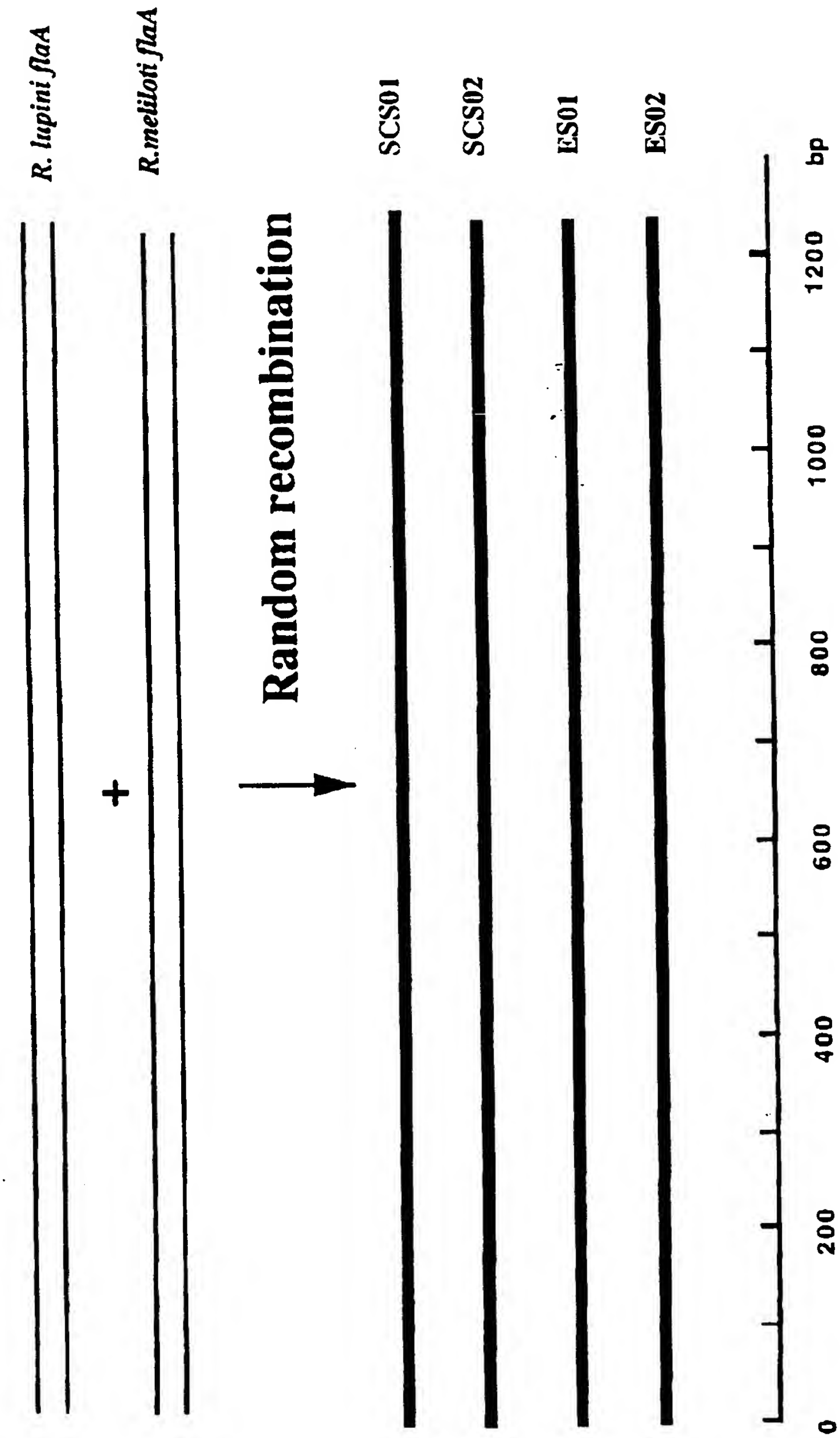
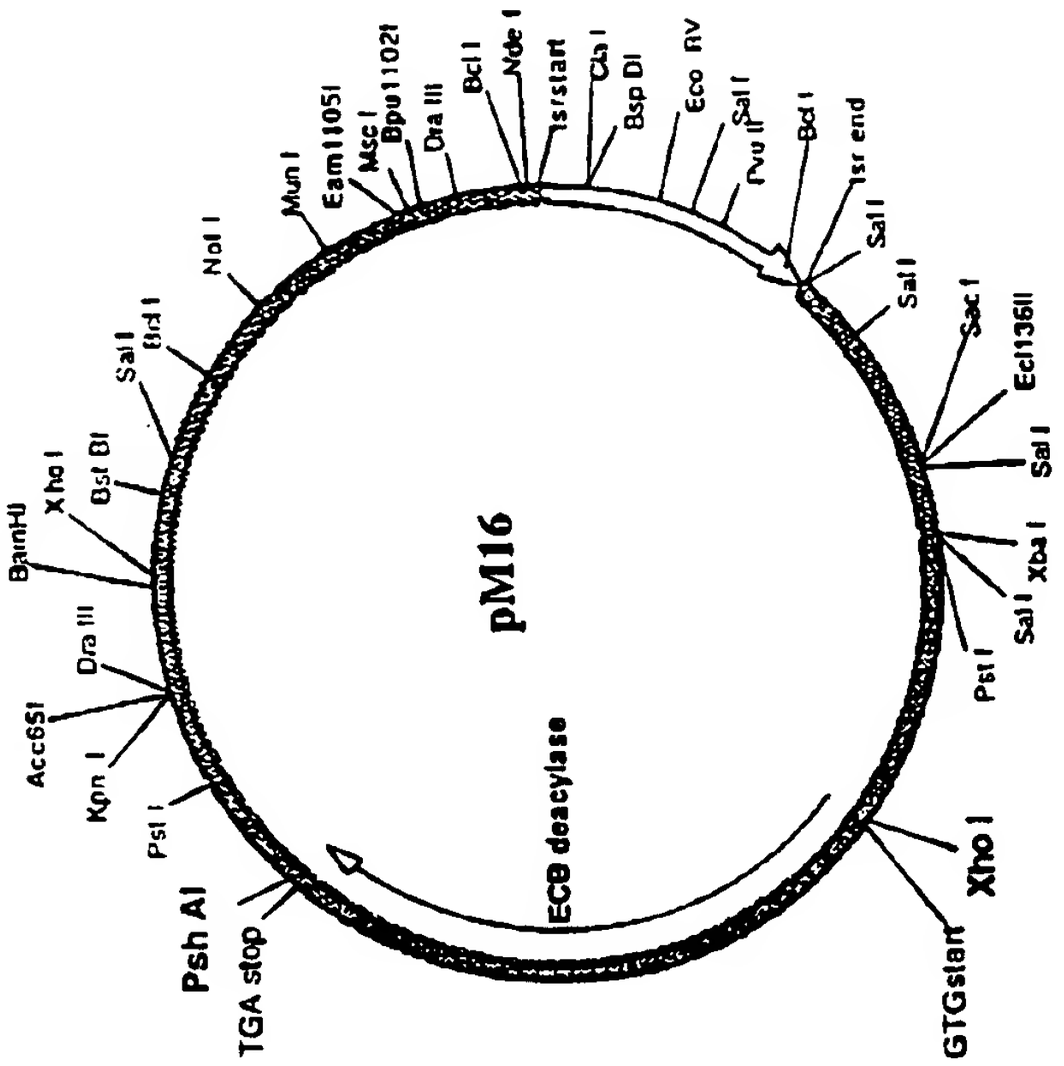

R. lupini flaA
R.meliloti flaA
+
Random recombination
SCS01
SCS02
ES01
ES02
bp
600
800
1000
1200
0
200
400

Fig. 8

(B)


pM16
ECB deacylase
Psh AI
TGA stop
GTGstart
Xho I
Pst I
Sal I
Xba I
Sal I
Sal I
Sac I
Ecl136II
Sal I
Sal I
Bcl I
Isr end
Pvu II
Sal I
Eco RV
Cla I
Bsp DI
Isrstart
Nde I
Bcl I
Dra III
Bpu1102I
Msc I
Eam1105I
Mun I
Not I
Sal I
Bcl I
Bst BI
Xho I
BamHI
Dra III
Acc65I
Kpn I
Psl I

(A)

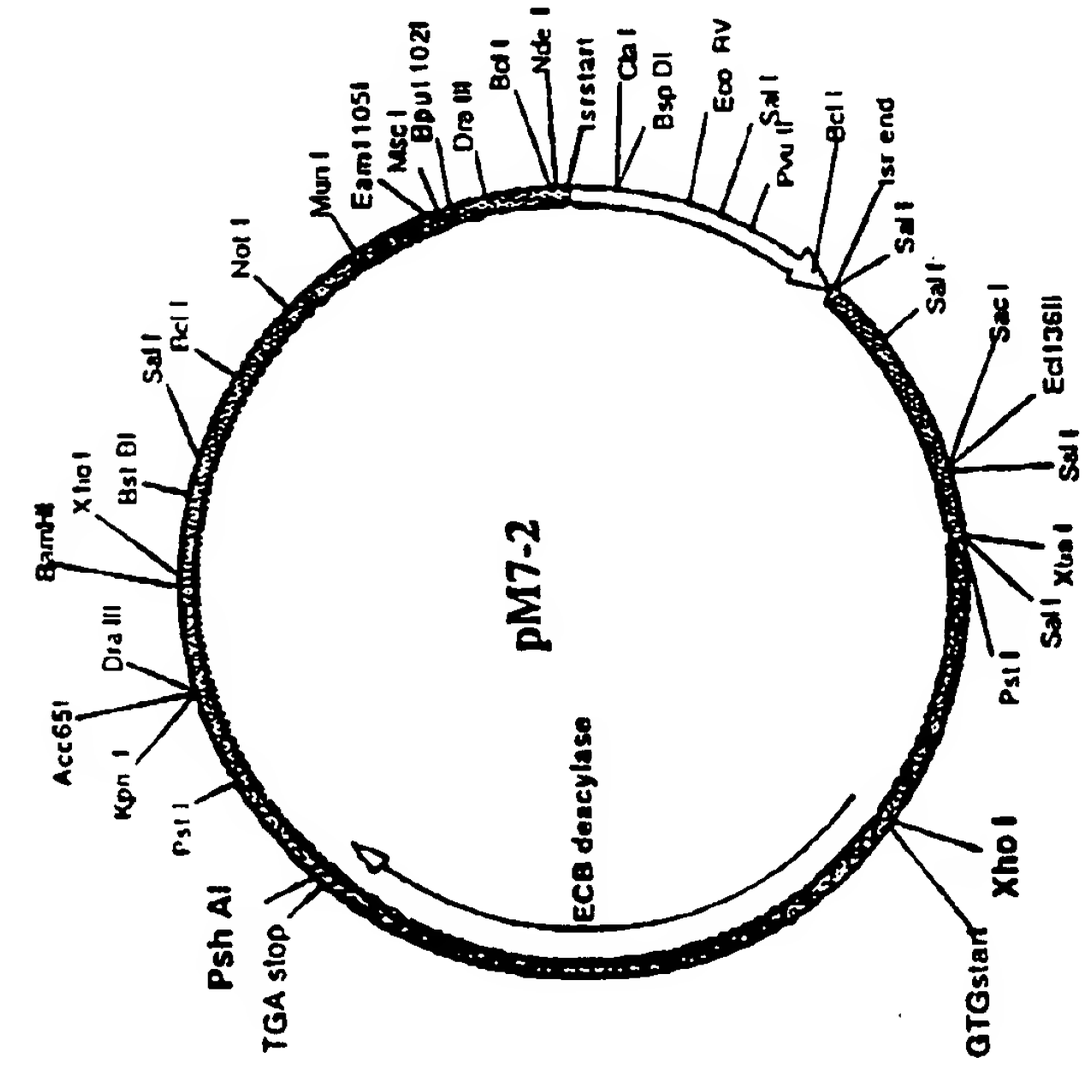

pM7-2
ECB deacylase
Psh AI
TGA stop
GTGstart
Xho I
Pst I
Sal I
Xba I
Sal I
Sal I
Sac I
Ecl136II
Sal I
Sal I
Bcl I
Isr end
Pvu II
Sal I
Eco RV
Cla I
Bsp DI
Isrstart
Nde I
Bcl I
Dra III
Bpu1102I
Msc I
Eam1105I
Mun I
Not I
Sal I
Bcl I
Bst BI
Xho I
BamHI
Dra III
Acc65I
Kpn I
Psl I

Fig. 9

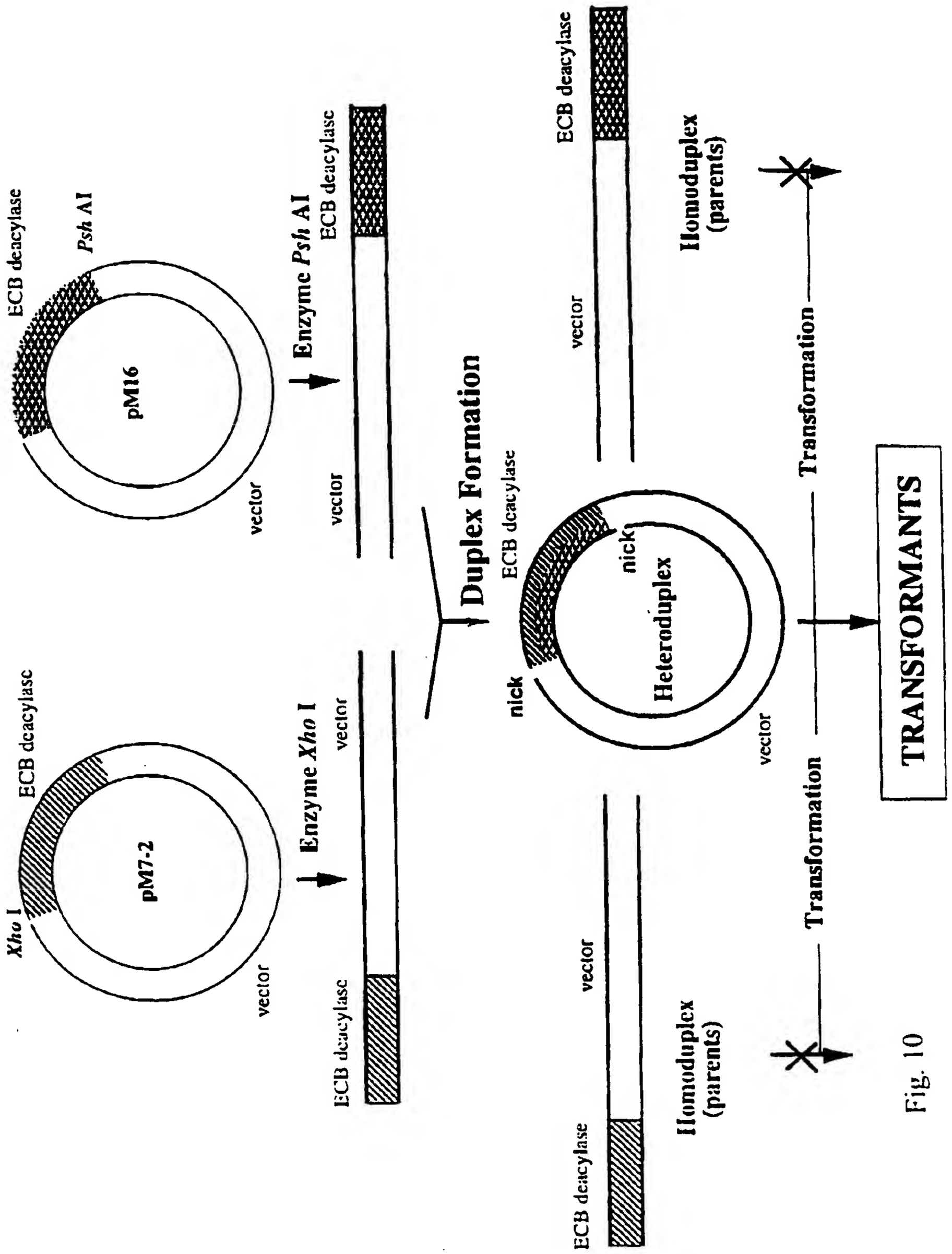

ECB deacylase
Psh AI
pM16
vector
Enzyme Psh AI
ECB deacylase
vector
ECB deacylase
vector
Homoduplex
(parents)
Enzyme Xho I
ECB deacylase
pM7-2
Xho I
vector
vector
Duplex Formation
ECB deacylase
nick
nick
Heteroduplex
vector
ECB deacylase
vector
Homoduplex
(parents)
Transformation
Transformation
TRANSFORMANTS

Fig. 10

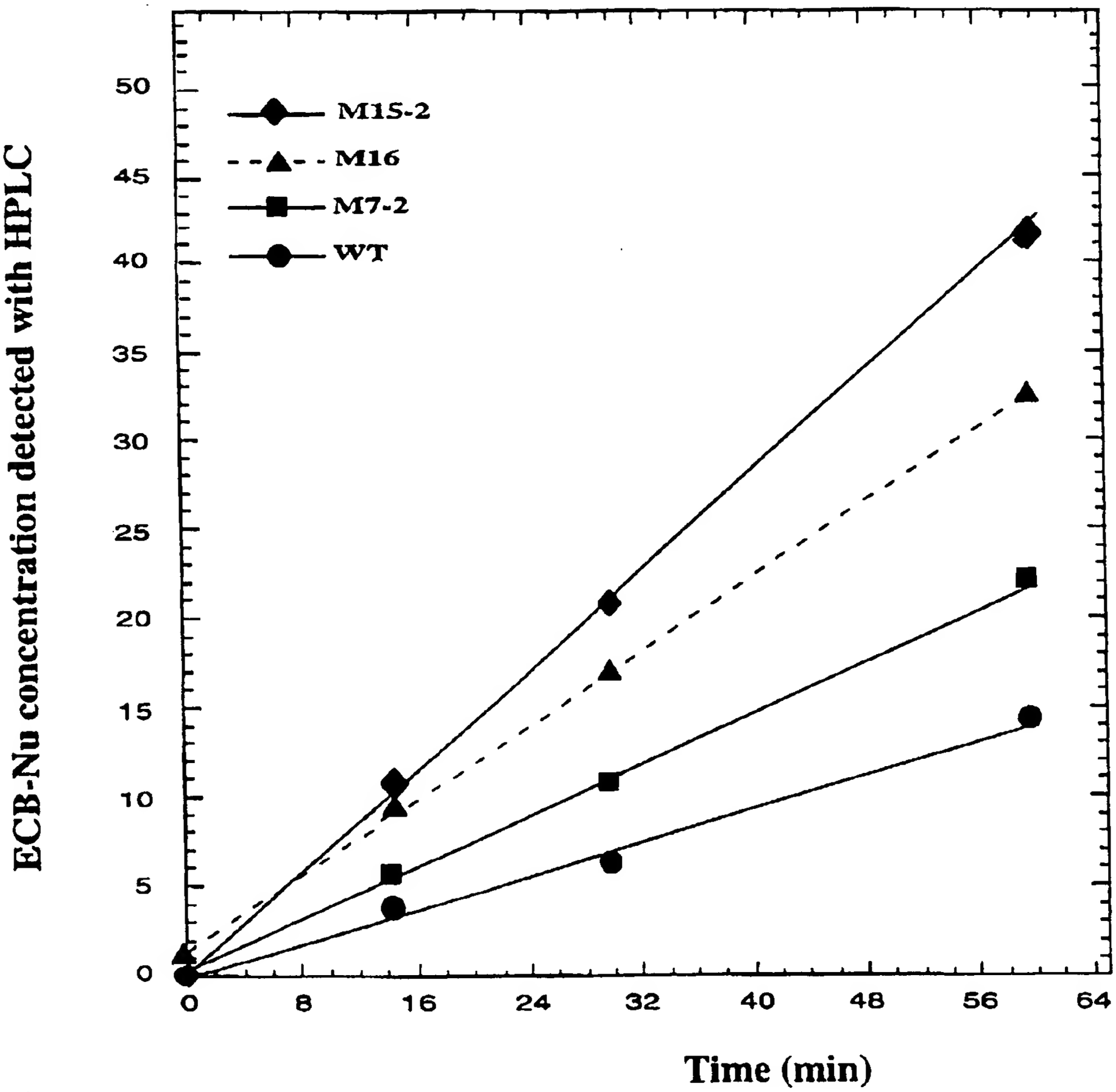

ECB-Nu concentration detected with HPLC
Time (min)
M15-2
M16
M7-2
WT
0
5
10
15
20
25
30
35
40
45
50
0
8
16
24
32
40
48
56
64

Fig. 11

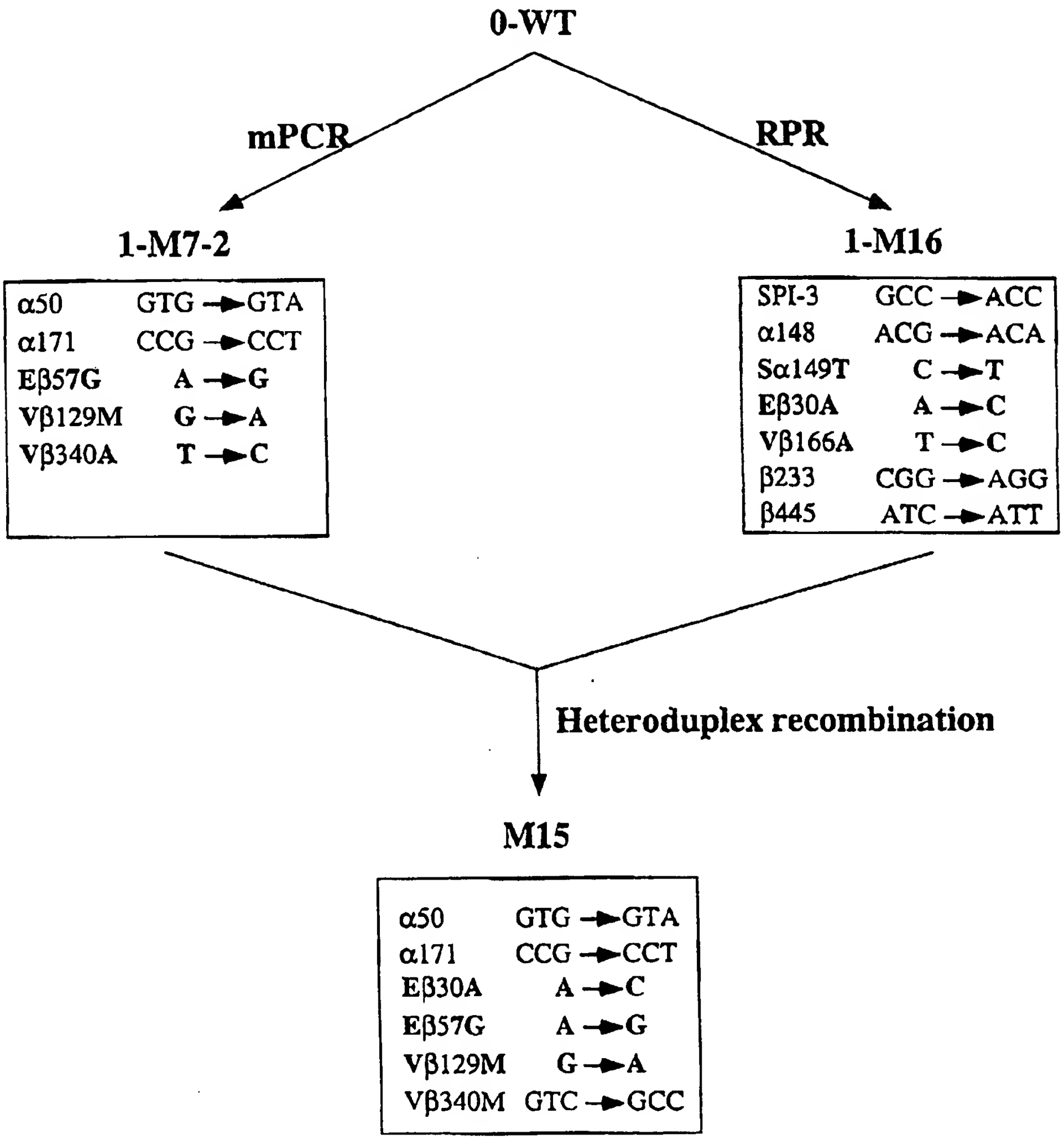

0-WT
mPCR
RPR
1-M7-2
α50 GTG → GTA
α171 CCG → CCT
Eβ57G A → G
Vβ129M G → A
Vβ340A T → C
1-M16
SPI-3 GCC → ACC
α148 ACG → ACA
Sα149T C → T
Eβ30A A → C
Vβ166A T → C
β233 CGG → AGG
β445 ATC → ATT
Heteroduplex recombination
M15
α50 GTG → GTA
α171 CCG → CCT
Eβ30A A → C
Eβ57G A → G
Vβ129M G → A
Vβ340M GTC → GCC

Fig. 12

```
1     CTGCAGCGTGCCCAGCTGTTCGTGGTGGTGATCGCGGCCGCGCTGGCCGCCGTCGCGGTC
61    GCCGCCGCCGGGCCGATCGAGTTCGTCGCCTTCGTCGTGCCGCAGATCGCCCTGCGGCTC
121   TGCGGCGGCAGCCGGCCGCCCCTGCTCGCCTCGGCGATGCTCGGCGCGCTGCTGGTGGTC
181   GGCGCCGACCTGGTCGCTCAGATCGTGGTGGCGCCGAAGGAGCTGCCGGTCGGCCTGCTC
241   ACCGCGATGATCGGCACCCCGTACCTGCTCTGGCTCCTGCTTCGGCGATCAAGAAAGGTG
301   AGCGGATGAACGCCCGCCTGCGTGGCGAGGGCCTGCACCTCGCGTACGGGGACCTGACCG
361   TGATCGACGGCCTCGACGTCGACGTGCACGACGGGCTGGTCACCACCATCATCGGGCCCA
421   ACGGGTGCGGCAAGTCGACGCTGCTCAAGGCGCTCGGCCGGCTGCTGCGCCCGACCGGCG
481   GGCAGGTGCTGCTGGACGGCCGCCGCATCGACCGGACCCCCACCCGTGACGTGGCCCGGG
541   TGCTCGGCGTGCTGCCGCAGTCGCCCACCGCGCCCGAAGGGCTCACCGTCGCCGACCTGG
601   TGATGCGCGGCCGGCACCCGCACCAGACCTGGTTCCGGCAGTGGTCGCGCGACGACGAGG
661   ACCAGGTCGCCGACGCGCTGCGCTGGACCGACATGCTGGCGTACGCGGACCGCCCGGTGG
721   ACGCCCTCTCCGGCGGTCAGCGCCAGCGCGCCTGGATCAGCATGGCGCTGGCCCAGGGCA
781   CCGACCTGCTGCTGCTGGACGAGCCGACCACCTTCCTCGACCTGGCCCACCAGATCGACG
841   TGCTGGACCTGGTCCGCCGGCTGCACGCCGAGATGGGCCGGACCGTGGTGATGGTGCTGC
901   ACGACCTGAGCCTGGCCGCCCGGTACGCCGACCGGCTGATCGCGATGAAGGACGGCCGGA
961   TCGTGGCGAGCGGGGCGCCGGACGAGGTGCTCACCCCGGCGCTGCTGGAGTCGGTCTTCG
1021  GGCTGCGCGCGATGGTGGTGCCCGACCCGGCGACCGGCACCCCGCTGGTGATCCCCCTGC
1081  CGCGCCCCGCCACCTCGGTGCGGGCCTGAAATCGATGAGCGTGGTTGCTTCATCGGCCTG
1141  CCGAGCGATGAGAGTATGTGGGCGGTAGAGCGAGTCTCGAGGGGGAGATGCCGCCGTGAC
                                                              V   T

1201  GTCCTCGTACATGCGCCTGAAAGCAGCAGCGATCGCCTTCGGTGTGATCGTGGCGACCGC
        3   S   S   Y   M   R   L   K   A   A   A   I   A   F   G   V   I   V   A   T   A

1261  AGCCGTGCCGTCACCCGCTTCCGGCAGGGAACATGACGGCGGCTATGCGGCCCTGATCCG
       23   A   V   P   S   P   A   S   G   R   E   H   D   G   G   Y   A   A   L   I   R

1321  CCGGGCCTCGTACGGCGTCCCGCACATCACCGCCGACGACTTCGGGAGCCTCGGTTTCGG
       43   R   A   S   Y   G   V   P   H   I   T   A   D   D   F   G   S   L   G   F   G

1381  CGTCGGGTACGTGCAGGCCGAGGACAACATCTGCGTCATCGCCGAGAGCGTAGTGACGGC
       63   V   G   Y   V   Q   A   E   D   N   I   C   V   I   A   E   S   V   V   T   A

1441  CAACGGTGAGCGGTCGCGGTGGTTCGGTGCGACCGGGCCGGACGACGCCGATGTGCGCAG
```

Fig. 13A

```
 83   N  G  E  R  S  R  W  F  G  A  T  G  P  D  D  A  D  V  R  S

1501 CGACCTCTTCCACCGCAAGGCGATCGACGACCGCGTCGCCGAGCGGCTCCTCGAAGGGCC
103   D  L  F  H  R  K  A  I  D  D  R  V  A  E  R  L  L  E  G  P

1561 CCGCGACGGCGTGCGGGCGCCGTCGGACGACGTCCGGGACCAGATGCGCGGCTTCGTCGC
123   R  D  G  V  R  A  P  S  D  D  V  R  D  Q  M  R  G  F  V  A

1621 CGGCTACAACCACTTCCTACGCCGCACCGGCGTGCACCGCCTGACCGACCCGGCGTGCCG
143   G  Y  N  H  F  L  R  R  T  G  V  H  R  L  T  D  P  A  C  R

1681 CGGCAAGGCCTGGGTGCGCCCGCTCTCCGAGATCGATCTCTGGCGTACGTCGTGGGACAG
163   G  K  A  W  V  R  P  L  S  E  I  D  L  W  R  T  S  W  D  S

1741 CATGGTCCGGGCCGGTTCCGGGGCGCTGCTCGACGGCATCGTCGCCGCGACGCCACCTAC
183   M  V  R  A  G  S  G  A  L  L  D  G  I  V  A  A  T  P  P  T

1801 AGCCGCCGGGCCCGCGTCAGCCCCGGAGGCACCCGACGCCGCCGCGATCGCCGCCGCCCT
203   A  A  G  P  A  S  A  P  E  A  P  D  A  A  A  I  A  A  A  L

1861 CGACGGGACGAGCGCGGGCATCGGCAGCAACGCGTACGGCCTCGGCGCGCAGGCCACCGT
223   D  G  T  S  A  G  I  G  S  N  A  Y  G  L  G  A  Q  A  T  V

1921 GAACGGCAGCGGGATGGTGCTGGCCAACCCGCACTTCCCGTGGCAGGGCGCCGCACGCTT
243   N  G  S  G  M  V  L  A  N  P  H  F  P  W  Q  G  A  A  R  F

1981 CTACCGGATGCACCTCAAGGTGCCCGGCCGCTACGACGTCGAGGGCGCGGCGCTGATCGG
263   Y  R  M  H  L  K  V  P  G  R  Y  D  V  E  G  A  A  L  I  G

2041 CGACCCGATCATCGGGATCGGGCACAACCGCACGGTCGCCTGGAGCCACACCGTCTCCAC
283   D  P  I  I  G  I  G  H  N  R  T  V  A  W  S  H  T  V  S  T

2101 CGCCCGCCGGTTCGTGTGGCACCGCCTGAGCCTCGTGCCCGGCGACCCCACCTCCTATTA
303   A  R  R  F  V  W  H  R  L  S  L  V  P  G  D  P  T  S  Y  Y
```

Fig. 13B

```
2161 CGTCGACGGCCGGCCCGAGCGGATGCGCGCCCGCACGGTCACGGTCCAGACCGGCAGCGG
323   V  D  G  R  P  E  R  M  R  A  R  T  V  T  V  Q  T  G  S  G

2221 CCCGGTCAGCCGCACCTTCCACGACACCCGCTACGGCCCGGTGGCCGTGATGCCGGGCAC
343   P  V  S  R  T  F  H  D  T  R  Y  G  P  V  A  V  M  P  G  T

2281 CTTCGACTGGACGCCGGCCACCGCGTACGCCATCACCGACGTCAACGCGGGCAACAACCG
363   F  D  W  T  P  A  T  A  Y  A  I  T  D  V  N  A  G  N  N  R

2341 CGCCTTCGACGGGTGGCTGCGGATGGGCCAGGCCAAGGACGTCCGGGCGCTCAAGGCGGT
383   A  F  D  G  W  L  R  M  G  Q  A  K  D  V  R  A  L  K  A  V

2401 CCTCGACCGGCACCAGTTCCTGCCCTGGGTCAACGTGATCGCCGCCGACGCGCGGGGCGA
403   L  D  R  H  Q  F  L  P  W  V  N  V  I  A  A  D  A  R  G  E

2461 GGCCCTCTACGGCGATCATTCGGTCGTCCCCCGGGTGACCGGCGCGCTCGCTGCCGCCTG
423   A  L  Y  G  D  H  S  V  V  P  R  V  T  G  A  L  A  A  A  C

2521 CATCCCGGCGCCGTTCCAGCCGCTCTACGCCTCCAGCGGCCAGGCGGTCCTGGACGGTTC
443   I  P  A  P  F  Q  P  L  Y  A  S  S  G  Q  A  V  L  D  G  S

2581 CCGGTCGGACTGCGCGCTCGGCGCCGACCCCGACGCCGCGGTCCCGGGCATTCTCGGCCC
463   R  S  D  C  A  L  G  A  D  P  D  A  A  V  P  G  I  L  G  P

2641 GGCGAGCCTGCCGGTGCGGTTCCGCGACGACTACGTCACCAACTCCAACGACAGTCACTG
483   A  S  L  P  V  R  F  R  D  D  Y  V  T  N  S  N  D  S  H  W

2701 GCTGGCCAGCCCGGCCGCCCCGCTGGAAGGCTTCCCGCGGATCCTCGGCAACGAACGCAC
503   L  A  S  P  A  A  P  L  E  G  F  P  R  I  L  G  N  E  R  T

2761 CCCGCGCAGCCTGCGCACCCGGCTCGGGCTGGACCAGATCCAGCAGCGCCTCGCCGGCAC
523   P  R  S  L  R  T  R  L  G  L  D  Q  I  Q  Q  R  L  A  G  T

2821 GGACGGTCTGCCCGGCAAGGGCTTCACCACCGCCCGGCTCTGGCAGGTCATGTTCGGCAA
543   D  G  L  P  G  K  G  F  T  T  A  R  L  W  Q  V  M  F  G  N
```

Fig. 13C

```
2881 CCGGATGCACGGCGCCGAACTCGCCCGCGACGACCTGGTCGCGCTCTGCCGCCGCCAGCC
563   R  M  H  G  A  E  L  A  R  D  D  L  V  A  L  C  R  R  Q  P

2941 GACCGCGACCGCCTCGAACGGCGCGATCGTCGACCTCACCGCGGCCTGCACGGCGCTGTC
583   T  A  T  A  S  N  G  A  I  V  D  L  T  A  A  C  T  A  L  S

3001 CCGCTTCGATGAGCGTGCCGACCTGGACAGCCGGGGCGCGCACCTGTTCACCGAGTTCGC
603   R  F  D  E  R  A  D  L  D  S  R  G  A  H  L  F  T  E  F  A

3061 CCTCGCGGGCGGAATCAGGTTCGCCGACACCTTCGAGGTGACCGATCCGGTACGCACCCC
623   L  A  G  G  I  R  F  A  D  T  F  E  V  T  D  P  V  R  T  P

3121 GCGCCGTCTGAACACCACGGATCCGCGGGTACGGACGGCGCTCGCCGACGCCGTGCAACG
643   R  R  L  N  T  T  D  P  R  V  R  T  A  L  A  D  A  V  Q  R

3181 GCTCGCCGGCATCCCCCTCGACGCGAAGCTGGGAGACATCCACACCGACAGCCGCGGCGA
663   L  A  G  I  P  L  D  A  K  L  G  D  I  H  T  D  S  R  G  E

3241 ACGGCGCATCCCCATCCACGGTGGCCGCGGGGAAGCAGGCACCTTCAACGTGATCACCAA
683   R  R  I  P  I  H  G  G  R  G  E  A  G  T  F  N  V  I  T  N

3301 CCCGCTCGTGCCGGGCGTGGGATACCCGCAGGTCGTCCACGGAACATCGTTCGTGATGGC
703   P  L  V  P  G  V  G  Y  P  Q  V  V  H  G  T  S  F  V  M  A

3361 CGTCGAACTCGGCCCGCACGGCCCGTCGGGACGGCAGATCCTCACCTATGCGCAGTCGAC
723   V  E  L  G  P  H  G  P  S  G  R  Q  I  L  T  Y  A  Q  S  T

3421 GAACCCGAACTCACCCTGGTACGCCGACCAGACCGTGCTCTACTCGCGGAAGGGCTGGGA
743   N  P  N  S  P  W  Y  A  D  Q  T  V  L  Y  S  R  K  G  W  D

3481 CACCATCAAGTACACCGAGGCGCAGATCGCGGCCGACCCGAACCTGCGCGTCTACCGGGT
763   T  I  K  Y  T  E  A  Q  I  A  A  D  P  N  L  R  V  Y  R  V

3541 GGCACAGCGGGGACGCTGACCCACGTCACGCCGGCTCGGCCCGTGCGGGGGCGCAGGGCG
```

Fig. 13D

```
783   A  Q  R  G  R

3601  CCGATCGTCTCTGCATCGCCGGTCAGCCGGGGCCTGCGTCGACCGGCGGCGGCCGGTCGA
3661  CGCCCGCGTCCCGGCGCAGCGACTGGCTGAAGCGCCAGGCGTCGGCGGCCCGGGGCAGGT
3721  TGTTGAACATCACGTACGCCGGGCCGCCGTCGAGGATGCCGGCGAGGTGTGCCAGCTCGG
3781  CATCCGTGTACACATGCCGGGCGCCGGTGATGCCGTGCAGCCGGTAATAGGCCATCGGCG
3841  TCAGACTGCGGCGCAGGAACGGGTCGGCGGCGTGGGTCAGGTCCAGCTCCTGGCACAAGC
3901  CCTCGACCACCTCGTCCGGCCACGGGCCGCGCGGCTCCCACAACAGCCGGACACCGGCCG
3961  GCCGGCGCGCTCGGGCGCAGAACTCACGCAGTCGCGCGATGGCGGGTTCGGTCGGCCGGA
4021  AACTCGCCGGGCACTGCAG
```

Fig. 13E

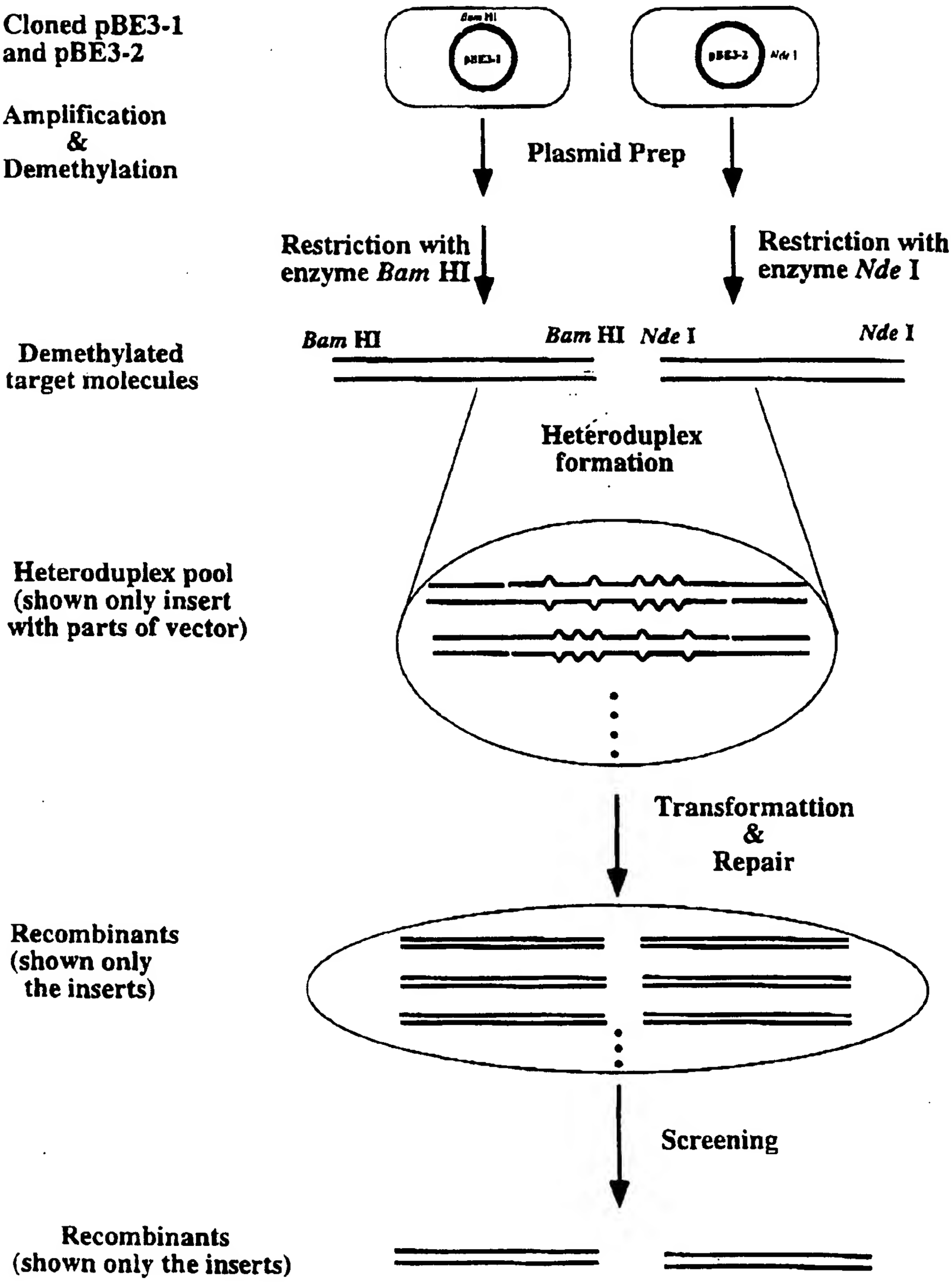

Cloned pBE3-1 and pBE3-2
Amplification & Demethylation
Plasmid Prep
Restriction with enzyme Bam HI
Restriction with enzyme Nde I
Demethylated target molecules
Bam HI
Bam HI
Nde I
Nde I
Heteroduplex formation
Heteroduplex pool (shown only insert with parts of vector)
Transformattion & Repair
Recombinants (shown only the inserts)
Screening
Recombinants (shown only the inserts)

Fig. 14

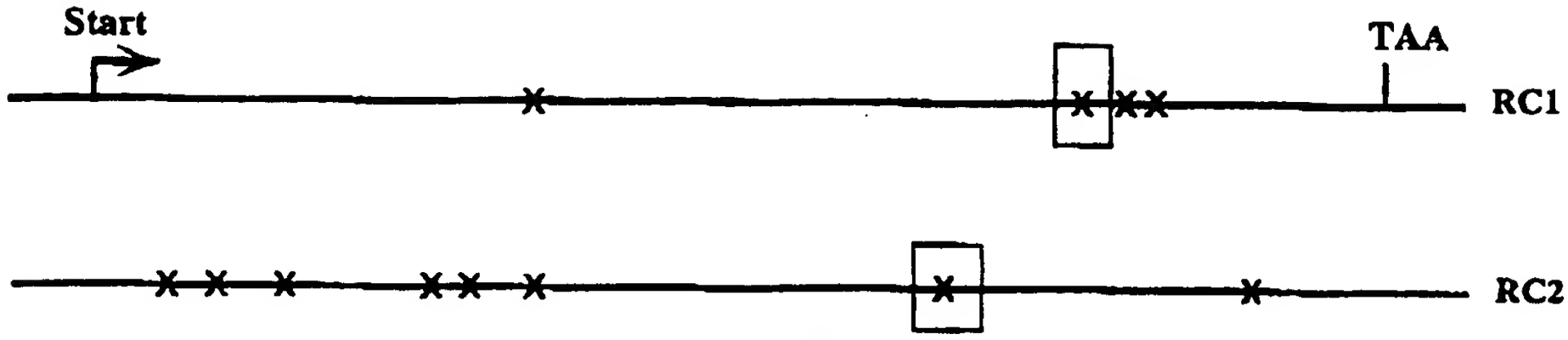

Start
TAA
RC1
RC2
HF/MR
Base position
484
520
598
731
745
780
995
1107
1141
1153
1189

Fig. 15

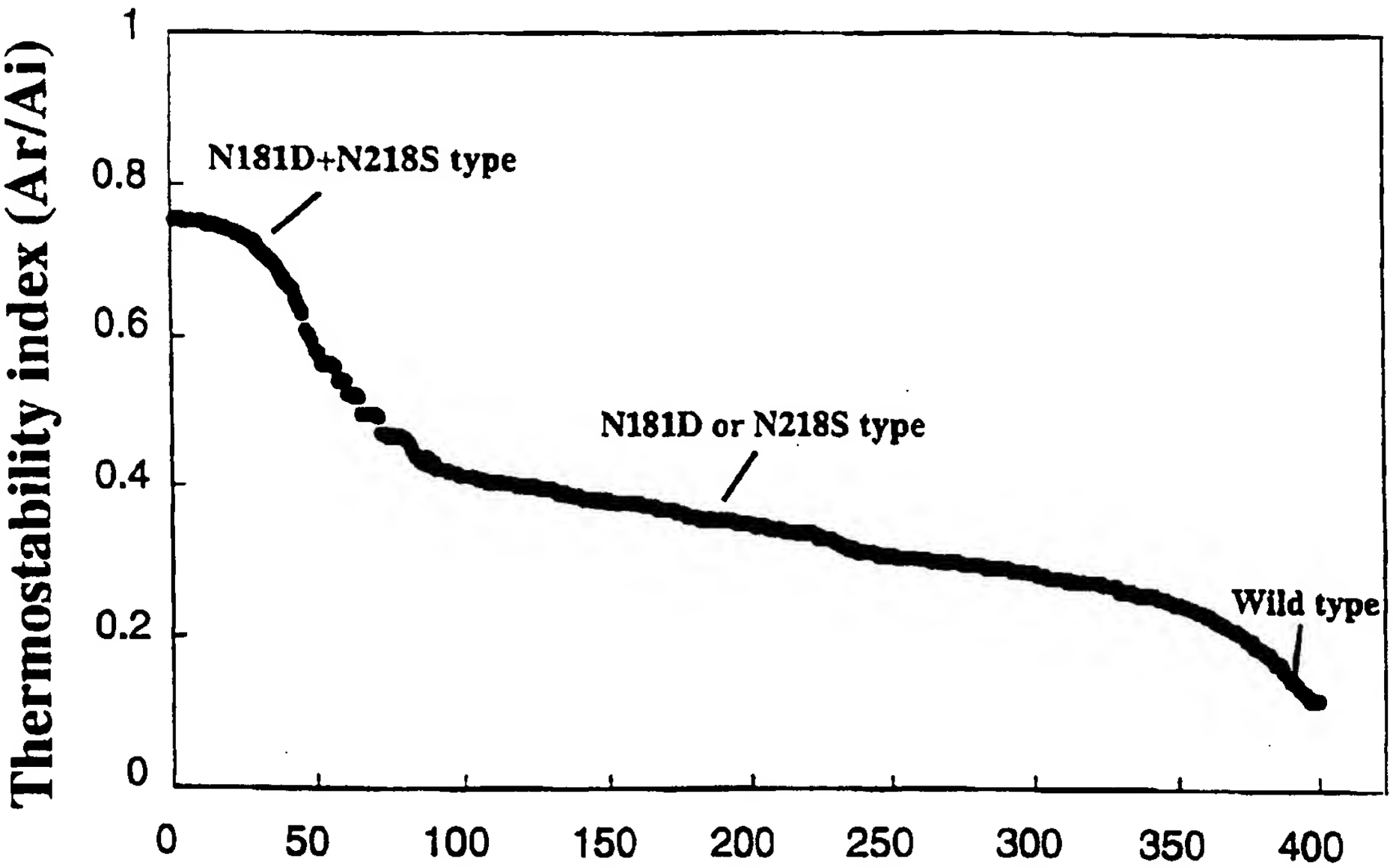

Thermostability index (Ar/Ai)
1
0.8
0.6
0.4
0.2
0
N181D+N218S type
N181D or N218S type
Wild type
0
50
100
150
200
250
300
350
400
Clones

Fig. 16

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/25698

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : C12Q 1/68; C12P 19/34
US CL :435/6, 91.1,91.2

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 91.1,91.2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category°	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,679,522 A (MODRICH et al) 21 October 1997 (10/21/97), see entire document, especially abstract.	1-34
Y	FANG, W.H. et al. Human Strand-specific Mismatch Repair Occurs by a Bidirectional Mechanism Similar to That of the Bacterial Reaction. J. of Bio. Chem. June 1993, Vol. 268, No. 16, pages 11838-11844, see entire document.	1-34
Y	US 5,556,750 A (MODRICH et al) 17 September 1996 (09/17/96), see entire document especially abstract.	1-34

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
26 MARCH 1999	26 APR 1999
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer JEFFREY SIEW Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)☆

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/25698

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS,STN,WPIDS,BIOSIS,MEDLINE,CANCERLIT,BIOTECHDS,LIFESCI,CAPLUS,EMBASE
search terms: amplification, heteroduplex, polynucleotides, repair,dna,pcr,vector